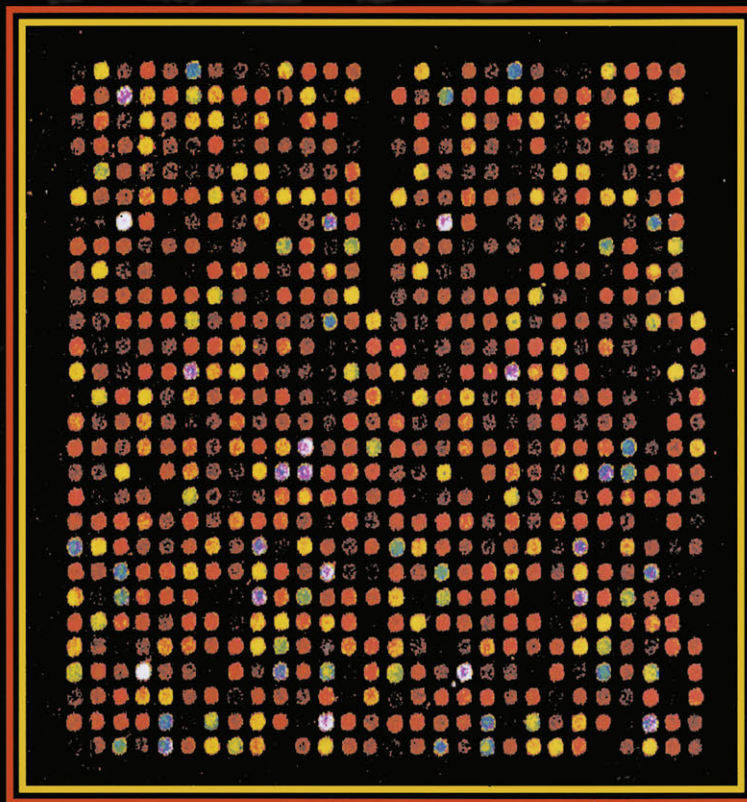

Molecular Techniques in Crop Improvement

Edited by

S. Mohan Jain, D.S. Brar and B.S. Ahloowalia



Springer-Science+Business Media, B.V.

Molecular Techniques in Crop Improvement

Molecular Techniques in Crop Improvement

Edited by

S. Mohan Jain

*International Atomic Energy Agency, FAO/IAEA Joint Division,
Vienna, Austria*

D.S. Brar

*International Rice Research Institute,
Los Banos, Philippines*

and

B.S. Ahloowalia

*Agriculture and Food Development Authority,
Dublin, Ireland*



SPRINGER-SCIENCE+BUSINESS MEDIA, B.V.

A C.I.P. Catalogue record for this book is available from the Library of Congress.

ISBN 978-90-481-5982-6 ISBN 978-94-017-2356-5 (eBook)
DOI 10.1007/978-94-017-2356-5

Printed on acid-free paper

All Rights Reserved

© 2002 Springer Science+Business Media Dordrecht

Originally published by Kluwer Academic Publishers in 2002

No part of the material protected by this copyright notice may be reproduced or utilized in any form or by any means, electronic or mechanical, including photocopying, recording or by any information storage and retrieval system, without written permission from the copyright owner.

TABLE OF CONTENTS

Preface	ix
1. Molecular genetics-plant breeder's perspective G.S. Khush	1
2. Molecular markers: principles and methodology P.K. Gupta, R.K. Varshney and M. Prasad	9
3. Molecular marker assisted breeding D.S. Brar	55
4. Microsatellites and molecular breeding: exploitation of microsatellite variability for the analysis of a monotonous genome P. Winter, B. Hüttel, K. Weising and G. Kahl	85
5. Development and application of molecular markers in conifers R. Schubert and G. Müller-Starck	139
6. DNA markers and heterosis P.K. Ranjekar, A.P. Davierwala and V.S. Gupta	161
7. Molecular markers and abiotic stresses I. Winicov	203
8. Molecular markers for flowering time genes in crop species D.A. Laurie and S. Griffiths	239
9. Molecular markers for the genetic analysis of apomixis M.J. Asins, M.R. Garcia, C. Ruiz and E.A. Carbonell	265

10. Use of molecular markers for fruit crop improvement A.R. McCaskill and J.J. Giovannoni	283
11. <i>In situ</i> hybridization in plants - methods and application J. Maluszynska	299
12. Molecular tools for improving coffee (<i>Coffea arabica</i> L.) resistance to parasites D. Fernandez and P. Lashermes	327
13. Construction and use of genetic maps in cereals M. Motto and P.A. Marsan	347
14. Gene expression under environmental stresses-molecular marker analysis M. Brosché, J.R. Gittins, H. Sävenstrand and Å. Strid	371
15. Random insertional mutagenesis in <i>Arabidopsis</i> T. Ito and K. Shinozaki	409
16. Engineering the chloroplast genome to confer stress tolerance and production of pharmaceutical proteins H. Daniell	427
17. Identification of strawberry flavour related genes by the use of DNA microarrays A. Aharoni and A.P. O'Connell	453
18. Gene targeting in plants S. Kumar and M. Fladung	481
19. Plant DNA methylation and gene expression M. Bellucci, F. Paolocci, F. Damiani and S. Arcioni	501
20. QTL mapping in crop plants S.B. Andersen and A.M. Torp	541

	vii
21. QTLs for rooth growth and drought resistance in rice	563
A. Price	
22. QTL mapping for forage traits	585
Th. Lübberstedt	
23. QTLs in bread making quality: a review	601
G. Charmet and C. Groos	

Preface

Plant breeding aims at the genetic enhancement of crops through the application of principles of Mendelian Genetics and modern tools and techniques of cell and molecular biology. Many breeding programs focus on the improvement of traits such as high yield, multiple resistance to major diseases, insect pests and tolerance to abiotic stresses and improved quality. The improved varieties must also fit into the crop rotation systems of different eco-agricultural regions for the production of feed, fiber, food, and industrial products. The value of new plant varieties in increasing food production has been demonstrated time and again, and perhaps the best of all, in the "Green revolution", which was based on the semi-dwarf rice and wheat varieties in Asia. Recent advances in molecular genetics have opened new opportunities to speed up plant breeding. Molecular markers have become important tools in the hands of plant breeders for enhancing the selection efficiency for various agronomic traits. Plant breeding has in fact entered in an era of genomics. The isolation, cloning and moving of genes from diverse biological sources into plant genomes holds promise to broaden the gene pool of crops and tailor plant varieties for specific traits that determine yield, quality and resistance to biotic and abiotic stresses.

This book is a sequel to our previous one on "Somaclonal Variation and Induced Mutations in Crop Improvement." The purpose of the present book is to bring together information on the new tools and techniques of molecular genetics in crop improvement. These tools include molecular markers and techniques such as RFLP, RAPD, AFLP, and PCR amplified DNA sequences. The molecular tools allow detection of specific DNA fragments through successive generations, and thus confirm transmission of the selected traits and incorporated genes. Many authors have contributed to the latest information on the available molecular technologies and their use to achieve practical results in plant breeding. This book should help plant breeders to choose the appropriate techniques suited to the needs of their

respective breeding programs, and allow them to integrate molecular tools based on gene cloning, gene maps, marker-assisted selection, QTL mapping, and molecular cytogenetics.

The book covers topics related to a wide range of molecular techniques and their application in crop improvement, and highlights molecular genetics from the perspective of plant breeders. The topics include principles and methodology of molecular markers, construction and use of molecular genetic maps, marker assisted breeding, characterization of alien gene introgression through molecular cytogenetics, and application of molecular markers in heterosis, apomixis, and tolerance to abiotic stress. A section of book has been devoted to the role of molecular markers for the improvement of fruit crops, conifers and coffee. Other sections focus on QTL mapping for drought tolerance, forage traits and bread making quality, and cover topics such as molecular basis of gene expression under stress, role of DNA methylation, insertional mutagenesis, gene targeting, engineering of chloroplasts and use of microarray technology.

We are grateful to all the authors who contributed their knowledge, wisdom and time for writing various chapters and submitted their manuscripts on schedule. We sincerely thank all our colleagues and friends who reviewed the manuscripts, and provided critical comments. We also take this opportunity to express our gratitude to Mr. Jacco Flipsen, Kluwer Academic Publishers, for his help in the successful completion of this project.

S. Mohan Jain

D.S. Brar

B.S. Ahloowalia

1

MOLECULAR GENETICS - PLANT BREEDER'S PERSPECTIVE

GURDEV S. KHUSH

International Rice Research Institute

Los Baños, Laguna, Philippines

Plant breeding is the art and science of changing and improving the heredity of plants to develop desired products with new genetic properties. The science of plant breeding consists of two phases, the evolutionary phase, which aims at creating or enlarging genetic variability, and the evaluation phase, which aims at selecting desirable genotypes from the variable populations. During the last four decades, major gains have been made in increasing productivity of major food crops worldwide. As an example, world rice production has more than doubled from 257 million tones in 1966 to 596 million tones in 1999. These gains in crop productivity have been achieved mainly through the application of principles of Mendelian Genetics and conventional plant breeding methods. However, to meet the growing need of ever increasing human population, the food grain production must increase by 50% in 2025. International Food Policy Research Institute (IFPRI, 1997) has estimated that maize production must increase by 80%, wheat by 60%, and rice by 40% during the next 25 years. There are no more suitable lands available for expansion of agriculture. Also, intensive agricultural systems have raised concerns about degradation of natural resources and environment deterioration. Therefore, we have to increase the food grain production from less land, less chemicals, less water, and less labor.

To achieve the goals of increased and sustainable crop production, there is an urgent need to develop varieties with higher yield potential and with greater yield stability. In the past, many approaches have been used to breed improved crop varieties. In fact, plant breeders have always used the best technology available for developing new and productive varieties with a combination of desirable

traits. Recent breakthroughs in cellular and molecular biology particularly in molecular markers and recombinant DNA technologies have provided new tools that can increase the efficiency of conventional plant breeding methods and allow several unconventional approaches for crop improvement. Some of the breakthroughs in molecular genetics include construction of dense molecular maps with several kinds of molecular markers (RFLP, RAPD, AFLP, SSR, etc.), availability of YAC, BAC, cDNA and EST libraries, discovery of extensive synteny across genomes and huge database on genomic sequences. A large number of genetic resources such as deletion mutants, insertional mutants, retrotransposon, and transposon-induced mutants are available for functional genomics. These can be used to identify useful genes for crop improvement. These developments provide most welcome additions to the tool box of plant breeders to solve plant breeding problems. These tools are useful for increasing the yield potential, developing cultivars with multiple resistance to biotic and abiotic stresses, improving nutritional quality of crops and in enhancing the efficiency of conventional plant breeding methods. Advances in molecular marker technology particularly the development of PCR based markers have enabled the application of marker assisted selection (MAS) in breeding programs to improve agronomic traits controlled by major genes as well as quantitative trait loci (QTL). Molecular markers are also useful for efficient management of germplasm and precise characterization of pathogen populations. Similar advances in gene isolation, identification of novel promoters, modification of gene expression and genetic engineering techniques have enabled production of transgenic crops with new genetic properties.

ENHANCING EFFICIENCY OF EVALUATION PHASE

Molecular markers and management of genetic resources

Genetic variability is the pre-requisite for any plant breeding program. Primitive cultivars, land races, elite breeding lines and wild relatives of crops comprise major component of genetic resources. A large number of germplasm collections of food, fiber, and horticultural crops are maintained in gene banks at various research centres. For example genetic resources of rice include 100,000 accessions. However, before these

resources can be exploited they must be systematically evaluated to assess genetic diversity. Classical approaches to detect genetic diversity are based on morphological traits, and isozyme polymorphism. However, morphological traits are highly influenced by the environment and hence estimates of genetic diversity are not precise. Isozymes are few in number, hence are not very efficient to characterize genetic diversity in the germplasm. On the other hand, the PCR based markers are abundant simple and reliable. Therefore they have become markers of choice to analyze genetic diversity in the germplasm collections of several food crops including horticultural and tree species.

DNA fingerprinting is an important approach to identify duplicates in the germplasm collections. Core collection or a subset of germplasm that might comprise up to 10 percent of the total and represent genetic diversity of a large collection can be evaluated through molecular markers. Such well characterized collections would be easy to maintain and serve as a source material for use in breeding programs for introducing desired variability to develop improved varieties of crop plants.

Rapid and precise characterization of pathogen populations

The availability of nucleic acid probes followed by dot-blot-spot hybridization method has been an important advancement to detect the presence of pathogens in the infected germplasm and is used widely to detect presence of viruses in potato, sweet potato and other tuber and root crops. The method is reliable, precise and can be completed in few hours. These DNA probes are also useful for screening segregating populations for disease resistance and thus improve the selection efficiency. Such probes can be used for the detection of viral, bacterial and fungal pathogens in the segregating populations. Similarly, genetic structures of pathogen populations can be determined based on restriction fragment pattern or through hybridization with the probes. Based on genetic

variability in pathogen populations, gene deployment strategies can be designed to enhance efficiency of resistance breeding.

Improving selection efficiency for agronomic traits through molecular marker assisted selection (MAS)

Numerous genes of economic importance such as those for disease and insect resistance are transferred from one varietal background to another by plant breeders. Most of these genes behave in a dominant and recessive manner and require time consuming efforts to transfer. Sometime screening procedures are cumbersome and expensive, and require large field space. If such genes can be tagged by tight linkage with molecular markers, time and money can be saved in transferring these genes from one varietal background to another. The presence or absence of the associated molecular marker would indicate at the very early stage, the presence or absence of desired target gene. A molecular marker very closely linked to the target gene can act as a tag which can be used for indirect selection of the gene in a breeding program. During the last decade dense molecular maps have been constructed in several crop species such as, wheat, barley, maize, rice, tomato and potato, etc. Several genes of economic importance have been tagged with molecular markers in these crop species (Khush and Brar 1998).

Once molecular markers closely linked to desirable traits are identified, marker assisted selection can be performed in early generation segregating populations and at early stage of development. With MAS it is now possible for the breeder to conduct several round of selection in a year. In many programs, MAS is integrated into existing plant breeding programs thus allowing researchers to access, transfer and combine genes at a rate and with precision not previously possible.

Gene pyramiding for durable resistance to pests

Durability of resistance can be enhanced if two or three genes for resistance to the same pathogen or pest can be pyramided into the same cultivar. Pyramiding of genes through conventional plant breeding approaches is difficult, laborious and in most cases impossible. MAS was successfully employed for pyramiding four genes for resistance to bacterial blight of rice e.g. *Xa4*, *xa5*, *xa13* and *Xa21* (Huang et al., 1997). The pyramided lines show wider spectrum of resistance as compared with lines having single resistance gene (Sanchez et al., 2000).

Identification and transfer of QTL through MAS

Most of the yield traits are polygenically inherited and are strongly influenced by the environment. Therefore, determination of genotypic values from phenotypic expression is not precise and selection strategies must take into account low heritabilities. Breeders generally select for yield when uniform breeding lines are obtained. Up to now it has not been possible to select for individual QTL having positive effect on yield in segregating populations. Recently individual QTL for yield have been tagged with molecular markers in rice. QTLs for several traits in rice such as blast resistance, salinity tolerance, submergence tolerance and root characters (length, thickness, dry weight and root shoot ratio) have also been identified and tagged with linked molecular markers.

Trait improving QTLs can also be transferred from exotic germplasm to elite lines via MAS. In rice, two yield enhancing QTLs (*yld1*, *yld2*) linked to molecular markers have been identified in *Oryza rufipogon* (Xiao et al., 1996). Similarly in tomato novel lines with increased yield over the original elite variety have been produced that contain specific QTLs from the wild species *Lycopersicon hirsutum*. These results demonstrate that discovery and transfer of QTLs for complex traits such as yield can be facilitated via molecular marker technology.

Isolation of agronomically useful genes via molecular map based cloning

A number of molecular approaches are available to isolate genes of agronomic importance. Map based cloning is one such strategy. Densely populated molecular genetic maps and availability of a large number of bacterial artificial chromosome (BAC) or yeast artificial chromosome (YAC) libraries have facilitated isolation of agronomically important genes such as *Xa1*, *Xa21*, *Pib* genes in rice (Song et al., 1995; Yoshimura et al., 1998; Wang et al., 1999), *Pto*, *Cf-2*, *Mi* in tomato (Martin et al., 1993; Dixon et al., 1996). Such native genes once cloned can be selectively transferred through transformation into commercial cultivars for value addition.

The development of molecular genetic maps has ushered in the era of comparative genetics or the science that exploits the results of "comparative mapping" (Devos and Gale, 1997). A remarkable level of synteny has been observed between the genomes of a number of cereal species despite differences in chromosome number and genome size. Such conservation of marker synteny is also helpful in cloning orthologous genes. Candidate regions for QTLs in one species are also candidate regions in related species. This can facilitate the isolation of desirable genes such as those for drought tolerance from one species and transfer it into other species through transformation.

Gene discovery through functional genomics

While genetic improvement through breeding has played a significant role in food production, many limiting physical and biological problems remain unsolved particularly in the less productive agricultural areas. Future gains will depend to a large extent upon our ability to manipulate complex traits that contribute to increased yield or better tolerance to multiple biotic and abiotic stresses. These traits are often difficult to

manipulate in a breeding program. Recent advances in genomics offer new opportunities to identify genes through functional genomics and to understand how genes work together to produce a trait. Extensive data have become available on sequences of rice and *Arabidopsis* genomes. The sequences of many plant genes will be known in the next few years. The availability of high-throughput technologies such as microarrays or gene chips, has made it possible to use mutants, isogenic lines, and recombinant inbred populations to assess gene function at a genomic scale. Microarrays containing cDNAs from biotic and abiotic stressed plants are being produced to identify candidate genes involved in biotic and abiotic stress pathways. Discovery of such genes and their manipulation would be the major step towards producing better crop varieties.

Broadening gene pool of crops through insertion of novel cloned genes

Conventional breeding methods enable plant breeders to exploit genetic variability from primary and secondary gene pools. Molecular biology tools have made it possible to transfer genes into crop plants from diverse sources such as unrelated plant and animal species as well as microorganisms through genetic engineering. This has helped broaden the gene pool of crops. Transgenic plants have been produced in more than 100 plant species. Several useful genes for herbicide tolerance, insect resistance and virus protection have been introduced into crop plants. Notable examples include *Bt*, *CpTi*, *gna*, *aai*, *pin2* for insect resistance; coat protein genes, chitinase, RIP for disease control; *bar*, *aroA*, *ALS*, *tfda* for herbicide tolerance; *gpat*, *sod*, *Mtld*, *betA*, *codA*, *imptI* for tolerance to abiotic stresses and *psy*, *crtl*, *lcy* for nutritional quality. Transgenic crops carrying novel genes with new genetic properties were grown in more than 44 million hectares of land in 2000 (James, 2000). The area under transgenic crops is likely to increase as more and more novel genes are introduced into crop varieties. Thus, transgenes are excellent sources for value addition in crop improvement.

REFERENCES

- Devos, K. M. and M. D. Gale. 1997. Comparative genetics in the grasses. *Plant Mol Biol.* 35:3-15.
- Dixon, M. S., D. A. Jones, J. S. Keddie, C. M. Thomas, K. Harrison and J. D. G. Jones. 1996. The tomato *Cf-2* disease resistance locus comprises two functional genes encoding leucine-rich repeat proteins. *Cell.* 84:451-459.
- Huang, N., E. R. Angeles, J. Domingo, G. Magpantay, S. Singh, G. Zhang, N. Kumaravadivel, J. Bennett and G. S. Khush. 1997. Pyramiding of bacterial blight resistance genes in rice: marker-assisted selection using RFLP and PCR. *Theor Appl Genet.* 95:313-320.
- IFPRI. 1997. Food Policy Research Institute, Washington, D. C.
- James, C. 2000. Global status of commercialized transgenic crops. ISAAA Briefs No. 21: Preview: ISAAA: Ithaca, N. Y. pp. 15.
- Khush, G. S. and D. S. Brar. 1998. The application of biotechnology to rice. In: *Agricultural Biotechnology in International Development.* pp. 97-121. (eds. C. L. Ives and B. M. Bedford). CAB International, Wallingford, UK.
- Martin, J. B., S. H. Brommonschenkel, J. Chunwongse, A. Frary, N. W. Ganai, R. Spivey, T. Wu, E. D. Earle and S. D. Tanksley. 1993. Map-based cloning of a protein-kinase gene conferring disease resistance in tomato. *Science.* 262:1432-1436.
- Sanchez, A. C., D. S. Brar, N. Huang, Z. Li and G. S. Khush. 2000. Sequence tagged site marker-assisted selection for three bacterial blight resistance genes in rice. *Crop Sci.* 40:792-797.
- Song, W. Y., G. L. Wang, L. L. Chen. 1995. A receptor kinase like protein encoded by the rice disease resistance gene, *Xa21*. *Science.* 270:1804-1806.
- Wang, Z. X., M. Yano, U. Yamanouchi, M. Iwamoto, L. Monna, H. Hayasaka, Y. Katayose and T. Sasaki. 1999. The *Pib* gene for rice blast resistance belongs to the nucleotide binding and leucine-rich repeat class of plant disease resistance genes. *Plant Journal.* 19: 55-64.
- Xiao, J., S. Grandillo, S. N. Ahn, S. R. McCouch, S. D. Tanksley, J. Li and L. Yuang. 1996. Genes from wild rice improve yield. *Nature.* 384:223-224.
- Yoshimura, S., U. Yamanouchi, Y. Katayose, S. Toki, Z. X. Wang, I. Kono, N. Kurata, M. Yano, N. Iwata and T. Sasaki. 1998. Expression of *Xa1*, a bacterial blight resistance gene in rice, is induced by bacterial inoculation. *Proc Natl Acad Sci. USA.* 95:1663-1668.

2

MOLECULAR MARKERS: PRINCIPLES AND METHODOLOGY

PUSHPENDRA K. GUPTA, RAJEEV K. VARSHNEY* AND MANOJ PRASAD*

Department of Agricultural Botany, Ch. Charan Singh University, Meerut-250 004, India

**Present address: Institut für Pflanzengenetik und Kulturpflanzenforschung (IPK), Corrensstraße 3, D 06466 Gatersleben, Germany*

1. INTRODUCTION

Molecular markers are based on the polymorphism detected at the level of macromolecules within the cell, although more recently, the term has largely been used to describe the DNA markers only. These DNA markers can be unlimited in number and can prove very useful for a variety of purposes relevant to crop improvement. For instance, these markers have been utilized extensively for the preparation of saturated molecular maps (genetical and physical). Their association with genes/QTLs controlling the traits of economic importance has also been utilized in some cases for indirect marker assisted selection (MAS). Other uses of molecular markers include gene introgression through backcrossing, germplasm characterization, genetic diagnostics, characterization of transformants, study of genome organisation, phylogenetic analysis, etc. (Rafalaski *et al.*, 1996). However, the progress in adoption of molecular markers by the plant breeders for MAS has been slow, despite the fact that in sharp contrast to the transgenic approach, no questions of biosafety and bioethics have been raised against this technology.

The first molecular markers in the form of restriction fragment length polymorphism (RFLP) had become available as early as 1980 (Botstein *et al.*, 1980), which made the beginning of an era of molecular crop breeding. Randomly amplified polymorphic DNAs (RAPDs), which became available in late 1980s, were the second in a range of different types of molecular

markers that were developed during 1980s and 1990s. RFLPs along with RAPDs, thus provided the first generation molecular markers, which were extensively used for DNA fingerprinting and partly also for MAS. These first generation DNA marker systems, however, did not live up to the initial expectations. Consequently, during 1990s, second generation molecular markers involving SSRs (Simple Sequence Repeats), AFLPs (Amplified Fragment Length Polymorphisms), and a variety of their modified forms were developed and utilized for a variety of purposes (for reviews, see Mohan *et al.*, 1997; Gupta *et al.*, 1999b). However, in future we will be using the third generation molecular markers including ESTs (Expressed Sequence Tags) and SNPs (Single Nucleotide Polymorphisms) that became available in mid and late 1990s. These marker systems do not always involve the time consuming and expensive gel-based assays needed for all other marker systems. Thus, during the last two decades, a broad range of molecular markers have become available, which are being used in a variety of ways not only to supplement and expedite the conventional methods for crop improvement, but also for characterization and maintenance of plant genetic resources (PGRs), that are so vital for crop improvement programmes. A large number of acronyms for these molecular markers are in common usage now (Table 1).

Although each marker system is associated with some advantages and disadvantages, the choice of marker system is dictated to a large measure by the intended application, convenience and the cost involved. High throughput approaches have also been developed, thus making it possible to scale-up the use of some of these markers. This has also attracted the private sector, particularly in the developed world, towards the use of molecular markers for crop improvement. The major uses and the factors controlling the choice of molecular markers have been discussed in several other chapters of this book. In this chapter, only the basic principles and methods involved in the development and uses of these molecular markers will be discussed. Furthermore, while describing the different classes of molecular markers in this chapter, the first and second generation molecular markers will be described only briefly, since they have recently been discussed at length in several reviews (Gupta *et al.*, 1996; Mohan *et al.*, 1997; Gupta *et al.*, 1999b; Gupta and Varshney, 2000). The third generation molecular markers (ESTs and SNPs) will be discussed in a relatively greater detail, not only because these markers have not been so widely discussed earlier, but also because they will be the markers of choice in future for a range of plant systems.

Table 1. List of acronyms for different DNA markers along with references

AFLP	= amplified fragment length polymorphism (Vos <i>et al.</i> , 1995)
AMP-PCR	= anchored microsatellite primed PCR (Wolff <i>et al.</i> , 1995)
AP-PCR	= arbitrarily primed PCR (Welsh and McClelland, 1990)
ASA	= allele-specific amplification (Wu <i>et al.</i> , 1989)
ASSR	= anchored simple sequence repeat (Wu <i>et al.</i> , 1994)
CAPS	= cleaved amplified polymorphic sequence (Akopyanz <i>et al.</i> , 1992)
DAF	= DNA amplification fingerprint (Caetano-Anollles <i>et al.</i> , 1991)
DALP	= direct amplification of length polymorphism (Desmarais <i>et al.</i> , 1998)
DAMD-PCR	= direct amplification of microsatellite DNA by PCR (Heath <i>et al.</i> , 1993)
DFLP	= DNA fragment length polymorphism (Hongtrakul <i>et al.</i> , 1998)
dRAMP	= digested RAMP (Becker and Heun, 1995)
IFLP	= intron fragment length polymorphism (Hongtrakul <i>et al.</i> , 1998)
IM-PCR	= inter-microsatellite PCR (Zietkiewicz <i>et al.</i> , 1994)
IRAP	= inter-retrotransposon amplified polymorphism (Kalendar <i>et al.</i> , 1999a)
ISA	= inter-SSR amplification (Zietkiewicz <i>et al.</i> , 1994)
ISSR	= inter-simple sequence repeats (Zietkiewicz <i>et al.</i> , 1994)
MAAP	= multiple arbitrary amplicon profiling (Caetano-Anollles, 1994)
MP-PCR	= microsatellite-primed PCR (Meyer <i>et al.</i> , 1993)
OLA	= oligonucleotide ligation assay (Landegren <i>et al.</i> , 1988)
RAHM	= randomly amplified hybridizing microsatellites (Cifarelli <i>et al.</i> , 1995)
RAMPO	= randomly amplified microsatellite polymorphisms (Richardson <i>et al.</i> , 1995)
RAMP	= randomly amplified microsatellite polymorphism (Wu <i>et al.</i> , 1994)
RAMS	= randomly amplified microsatellites (Ender <i>et al.</i> , 1996)
RAPD	= random amplified polymorphic DNA (Williams <i>et al.</i> , 1990)
RBIP	= retrotransposon-based insertion polymorphism (Flavell <i>et al.</i> , 1998)
REMAP	= retrotransposon-microsatellite amplified polymorphism (Kalendar <i>et al.</i> , 1999a)
RFLP	= restriction fragment length polymorphism (Botstein <i>et al.</i> , 1980)
SAMPL	= selective amplification of microsatellite polymorphic loci (Morgante and Vogel, 1994)
SCAR	= sequence characterised amplified regions (Paran and Michelmore, 1993)
SNP	= single nucleotide polymorphism (Landegren <i>et al.</i> , 1988)
SPAR	= single primer amplification reactions (Gupta <i>et al.</i> , 1994)
S-SAP	= sequence-specific amplification polymorphism (Waugh <i>et al.</i> , 1997)
SSCP	= single strand conformation polymorphism (Hayashi, 1992)
SSLP	= simple sequence length polymorphism (Tautz, 1989)
SSR	= simple sequence repeat (Hearne <i>et al.</i> , 1992)
STAR	= sequence tagged amplified region (Rafalaski and Tingey, 1993)
STMS	= sequence-tagged microsatellite site (Beckmann and Soller, 1990)
STR	= short tandem repeat (Edwards <i>et al.</i> , 1991)
STS	= sequence- tagged- site (Olson <i>et al.</i> , 1989)
VNTR	= variable number of tandem repeats (Nakamura <i>et al.</i> , 1987)

2. TYPES OF MOLECULAR MARKERS

On the basis of the principles and methods employed, molecular markers can be broadly classified in the following four groups: (i) hybridization based markers, (ii) PCR based markers, (iii) molecular markers based on PCR followed by hybridization, and (iv) sequencing and DNA chip based markers (for a review, see Mohan *et al.*, 1997; Gupta *et al.*, 1999b). Each of these four types will be briefly described in this chapter. For more details, readers may like to consult our two recent reviews (Gupta *et al.*, 1999b; Gupta and Varshney, 2000) and other original works cited at appropriate places in this chapter.

2.1 Hybridization Based Molecular Markers

2.1.1 Restriction Fragment Length Polymorphisms (RFLPs)

RFLPs are based on polymorphisms arising due to base substitutions, insertions, deletions and translocations that might have occurred in the genomic DNA in the past. These changes lead to differences in the size of restriction fragments obtained due to digestion with specific restriction enzymes (Fig. 1). RFLPs detect differences only in a fraction of these fragments, which are related with each other through homology with a molecular probe used for hybridization. Following steps are involved: (i) digestion of genomic DNA with individual restriction enzymes, followed by electrophoresis on agarose gel (1%- 2%); (ii) transfer of DNA fragments from the gel to a filter following the technique of Southern blotting (Southern, 1975); (iii) hybridization of DNA fragments on the filter with one or more molecular probes individually (the probes need to be labelled radioactively with ^{32}P or non-radioactively with biotin/digoxigenin, before hybridization), and (iv) study of polymorphism in the hybridization patterns. The molecular probes used for RFLP can be either genomic DNA probes (derived from genomic DNA) or cDNA probes (derived from mRNA), which can be developed from genomic/ cDNA libraries in the laboratory or else can be procured from other laboratories. Heterologous probes derived from one species can also be used across related species and genera. Rarely, synthetic oligonucleotides may also be used as molecular probes. The level of polymorphism detected through RFLPs depends both on the restriction enzyme used for digestion and the probe used for hybridization.

One of the advantages of RFLPs is due to their co-dominant nature, which allows distinction between homozygotes and heterozygotes. Another advantage is due to the expected homology between the polymorphic

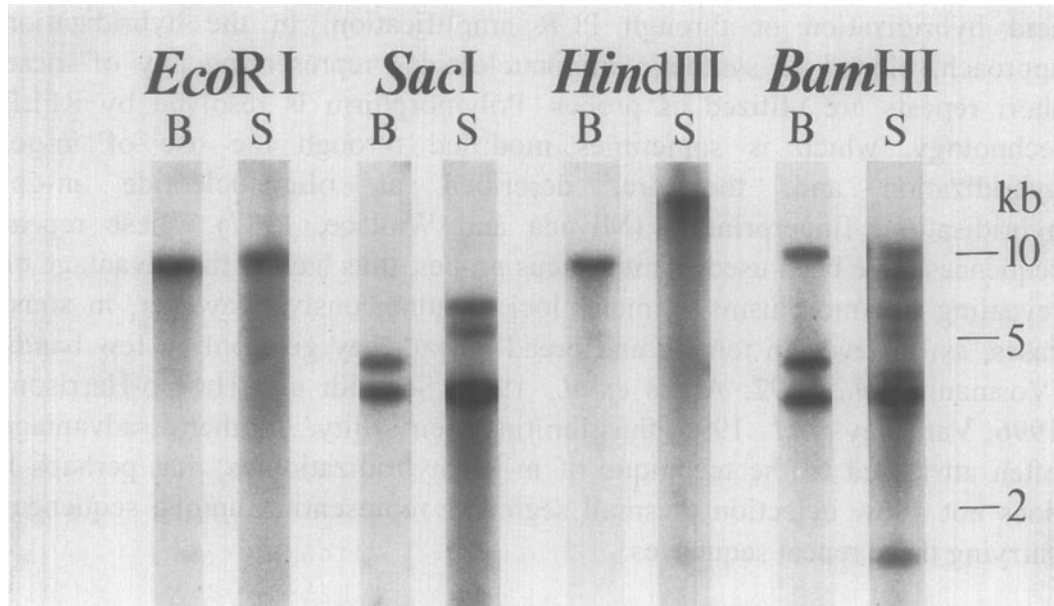


Figure 1. RFLPs of *H. bulbosum* (B) and *H. spontaneum* (S) DNA digested with *EcoRI*, *SacI*, *HindIII* and *BamHI* and probed with pTA71 (ribosomal DNA) (reproduced from Molnar *et al.*, 1989, with permission).

fragments scored, since they hybridize with the same probe (compare with RAPDs). RFLP analysis, however, is time consuming, labour intensive and is too slow for rapid evaluation of large segregating populations used in a commercial breeding programme (Gale *et al.*, 1995). An additional disadvantage of RFLP analysis is due to large quantities of genomic DNA (10 µg or more), needed for the preparation of Southern blots. Because of these limitations, and due to superiority and cost effective nature of PCR based marker systems, which became available, later, RFLPs are seldom, if ever used now for developing molecular markers for marker aided selection (MAS) and other similar uses.

2.1.2 Dispersed Repetitive DNA (drDNA) for DNA Fingerprinting

Nuclear genomes of all eukaryotes contain repeats of short sequence motifs (2-15 bp in length), dispersed throughout the genome, which have been described either as dispersed repetitive DNA (drDNA; Whitkus *et al.*, 1994),

or as Variable Number of Tandem Repeats (VNTRs; Nakamura *et al.*, 1987). These sequences mainly include minisatellites and microsatellites {The microsatellites have also been described as Simple Sequence Repeats (SSRs) or Short tandem repeats (STRs)} and can be detected, either through nucleic acid hybridization or through PCR amplification. In the hybridization approach, cloned or synthetic oligonucleotides representing any of these short repeats are utilized as probes. Polymorphism is resolved by RFLP technology, which is sometimes modified through the use of in-gel hybridization and, therefore, described as oligonucleotide in-gel hybridisation/ fingerprinting (Miyada and Wallace, 1987). These repeat sequences have been used as multilocus probes, thus having the advantage of revealing polymorphisms at many loci simultaneously. However, in some cases, as observed in tomato and bread wheat, they give only a few bands (Vosman *et al.*, 1992; Arens *et al.*, 1995; Schmidt and Heslop-Harrison, 1996; Varshney *et al.*, 1998) thus limiting their utility. Another disadvantage often attributed to the technique of in-gel hybridization is, that perhaps it does not allow detection of small segments representing unique sequences carrying these repeat sequences.

2.2 PCR BASED MOLECULAR MARKERS

PCR based molecular markers offer the potential to reduce time, effort and expense required for molecular marker technology. They are based on the use of a pair of primers (forward and reverse primers), which may be designed either on the basis of arbitrary sequences or on the basis of specific sequences flanking the DNA segment that needs to be amplified. Sometimes the same arbitrary or specific sequence may be used both for forward and reverse primers, so that the PCR based markers may be classified into two types, one using a single primer, and the other using a pair of primers.

2.2.1 Single Primer PCR Methods

2.2.1.1 Arbitrary Primed PCR (AP-PCR), Random Amplified Polymorphic DNA (RAPD) and DNA Amplification Fingerprinting (DAF)

The basic principle involved in these three similar techniques assumes the presence of same arbitrary or random sequence in inverse orientation within an amplifiable distance, so that the same sequence works as forward and

reverse primers at multiple loci. The techniques, therefore, have also been collectively termed as Multiple Arbitrary Amplicon Profiling (MAAP; Caetano-Anolles and Gresshoff, 1994) or Arbitrarily Amplified DNA (AAD; Caetano Anolles, 2001). The three techniques mainly differ in the length of the primer sequence, which is ~20 nucleotides long for AP-PCR (Welsh and McClelland, 1990), ~10 nucleotides long for RAPDs (Williams *et al.*, 1990) and ~6-8 nucleotides long for DAF (Caetano-Anolles *et al.*, 1991). AP-PCR and RAPDs are generally considered to be similar, although RAPDs became more popular, since, it gives relatively higher number of bands due to smaller size of primers, thus making it suitable for fingerprinting. Moreover, a large number of suitably designed random primers for RAPDs are commercially available, thus making RAPDs (often pronounced as 'rapids') more popular (Fig. 2).

Later it was realized, that with the reduction in the size of primer sequence, frequency of the occurrence of primer sequence in the genome will increase, leading to increase in the number of amplified products. Consequently, in DAF with 6-8 nucleotides long primers, the number of amplified products increases many fold (Fig. 3). This made DAF relatively more useful for DNA fingerprinting (for more details, see Caetano-Anolles *et al.*, 1991). In all the above three methods, if the sequence of primer binding sites at a locus mismatches the primer sequence, this will not allow binding of the primer(s) and no amplified product will be obtained, thus giving a plus/minus type of polymorphism. Consequently, these markers are generally dominant in nature, although the possibility of their rarely being co-dominant can not be ruled out. The co-dominant RAPD markers will be available, when fragments of different sizes may represent length variation in homologous sequences, which is not unlikely. In other cases, during AP-PCR/RAPD or DAF analysis, the fragments, that are scored to be similar on the basis of size, may sometimes exhibit sequence variation (SNP alleles exhibiting SSCP represent one such example, although the variation may be of a much higher magnitude). These anomalies are inherent in the technique, since homology of fragments is not tested using a probe as done in RFLP. Further, the use of these markers (particularly RAPDs) across the laboratories may not be reproducible due to variation in PCR conditions and different models of thermal cyclers used. Advantages of AP-PCR/RAPDs/DAF over RFLPs and other PCR based markers, however, include the following: (i) AP-PCR/RAPDs and DAF do not require labour intensive and expensive procedures like those needed in RFLP; (ii) no genomic/cDNA libraries are needed for development of probes, (iii) for each PCR reaction very small

quantity of genomic DNA is needed (~100 ng in AP-PCR/RAPD/DAF vs 10 µg in RFLP) and (iv) no sequence information is required for designing the primers (as needed in SSRs).

2.2.1.2 Amplification of drDNA/VNTRs and Retroposons

Using PCR, DNA polymorphism can also be studied at loci harbouring drDNA or retroposons (see above for drDNA). The technique relies on the presence of same repeat sequence (based on drDNA/retroposon) in inverse orientation, separated by an amplifiable distance within the genome, so that the inter-repeat sequences are amplified. Due to higher stringency of annealing, the reproducibility of this technique is known to be better than RAPDs, although the banding pattern is known to be influenced by all PCR parameters, especially the annealing temperature (Meyer *et al.*, 1993). In all the approaches under this category also, generally plus/minus type of polymorphism is scored, so that these markers are also generally dominant in nature. The amplified fragments obtained in this approach, can also be isolated and cloned to be used as molecular probes on the genomic DNA (Somers *et al.*, 1996). Like unique sequence probes often used in RFLPs, these drDNA or retroposon probes also reveal polymorphism and generate DNA fingerprinting patterns, which could be used for species differentiation and cultivar identification (Bebeli *et al.*, 1997). If tried, these markers should

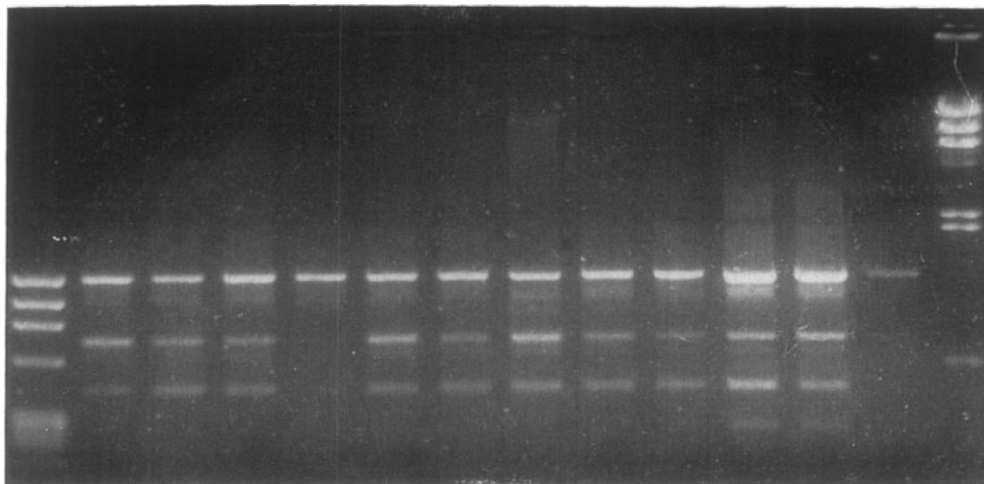


Figure 2. RAPD profiles in 12 rice genotypes obtained using a RAPD primer 10 nucleotides in length.

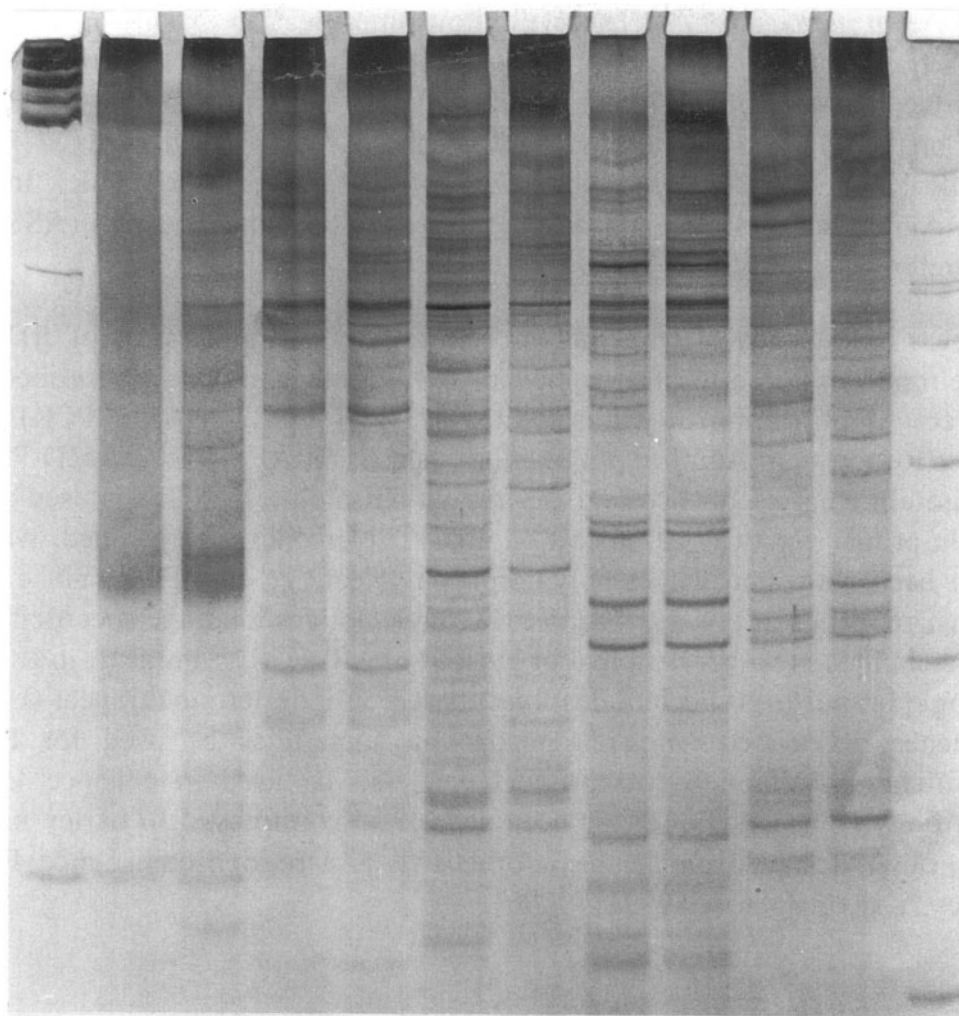


Figure 3. DAF profiles of two bread wheat genotypes (1, 2) obtained using five different mini-hairpin primers (two adjacent lanes for each primer).

also prove useful for molecular mapping, and for tagging of genes for economic traits. The clones obtained due to elution of the amplified fragments can also be converted into locus specific PCR-based markers, if primers are designed by sequencing these fragments.

There are at least three similar approaches (MP-PCR, DAMD-PCR, IRAP), where polymorphism associated with drDNA and/or retroposons can be examined. In one such strategy, described as Microsatellite Primed-PCR (MP-PCR), synthetic oligonucleotides representing di- or trinucleotide repeats have been used as single PCR primers for the amplification of discrete regions in genomic DNA. Sometimes, radiolabelled synthetic oligonucleotides representing di- or trinucleotide repeats, anchored through 2-4 nucleotides at one of the two ends, are used as single PCR primers

(Zietkiewicz *et al.*, 1994). This will allow amplification of only a subset of DNA fragments that are amplified by MP-PCR and may, therefore, lead to improvement in reproducibility, a problem that limits the use of MP-PCR. The method is known variously as Anchored Microsatellite-Primed PCR (AMP-PCR), Inter-Microsatellite PCR (IM-PCR), Inter-SSR (ISSR), Inter-SSR Amplification (ISA) and Anchored Simple Sequence Repeats (ASSRs) (for references, see Table 1).

Success in the study of DNA polymorphism due to amplification of drDNA or retroposons has also been achieved by two other methods, described as “Direct Amplification of Minisatellites DNA by PCR” (DAMD-PCR) and “Inter-Retroposon Amplified Polymorphism” (IRAP). In DAMD-PCR, minisatellite core sequence (e.g. Jeffrey’s minisatellite or M13) is used as a single primer for PCR amplification and PCR profiles are obtained, which have been shown to detect polymorphism (Heath *et al.*, 1993). Since the primer sequence in this approach is 12-15 nucleotides long, the specificity of this technique is higher than that of RAPDs or MP-PCR. Similarly, in IRAP, a single primer (~20 nucleotides long) based on long terminal repeat (LTR) sequence, associated with retroposon can sometimes be used for PCR amplification, although two primers can also be used to enhance locus specificity (Fig. 4a). IRAP has been successfully employed in barley using the primer designed on the basis of BARE-1, a retroposon isolated from barley (Kalendar *et al.*, 1999a; 1999b).

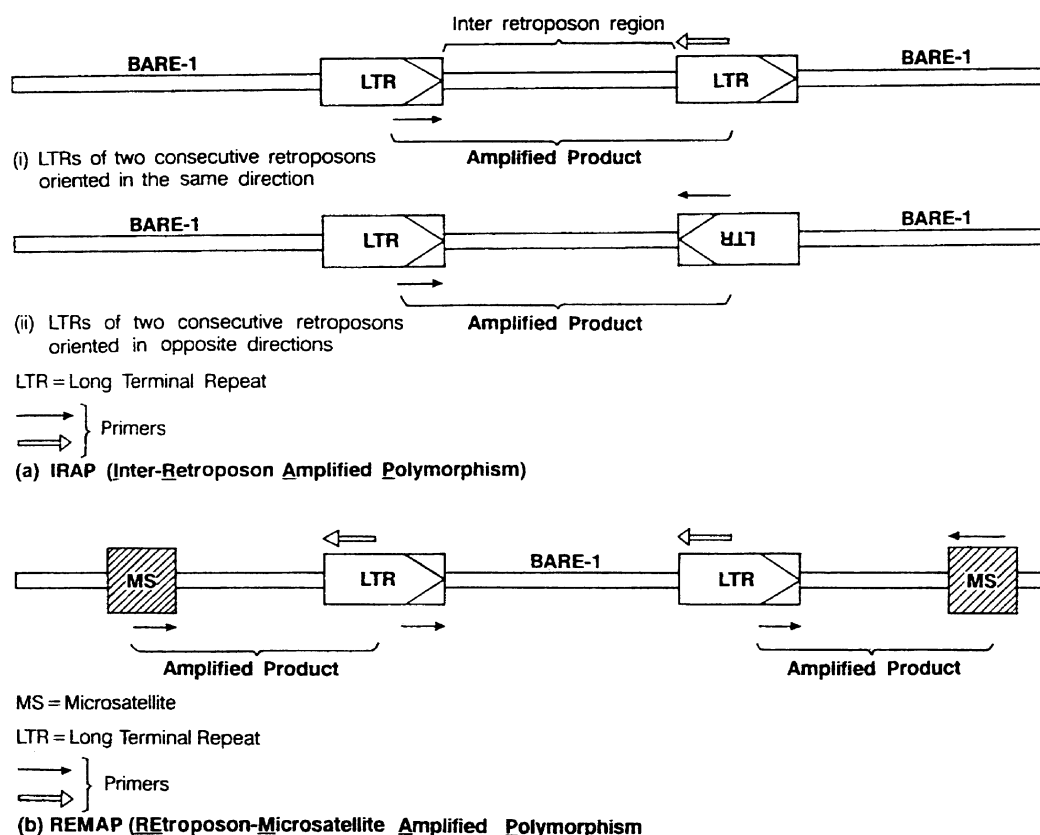


Figure 4. Diagrammatic representation of the principle involved in IRAP (a) and REMAP (b) molecular markers (modified from Kalendar *et al.*, 1999).

2.2.2 PCR Methods Using a Pair of Primers

2.2.2.1 Sequence-Tagged Sites (STSs) and Sequence Characterized Amplified Regions (SCARs)

A Sequence Tagged Site (STS) is a short unique sequence that identifies one or more specific loci, which can be amplified through PCR. Each STS is characterized by a pair of PCR primers, which are designed by partial sequencing of a RFLP probe (including genomic DNA and cDNA probes) representing a mapped low copy number DNA sequence (Fig. 5). STS primers can also be designed from the EST sequences available in databases (see later). Like STS primers, which identify RFLP loci, SCAR primers identify loci that are amplified in PCR based marker systems including RAPD/AP-PCR, DAF, or AFLP/SAMPL (see later for AFLP/SAMPL). According to Rafalaski and Tingey (1993), a more suitable acronym for SCARs is STARS (Sequence-Tagged Amplified Regions). A higher level of polymorphism may also be detected by restriction digestion of

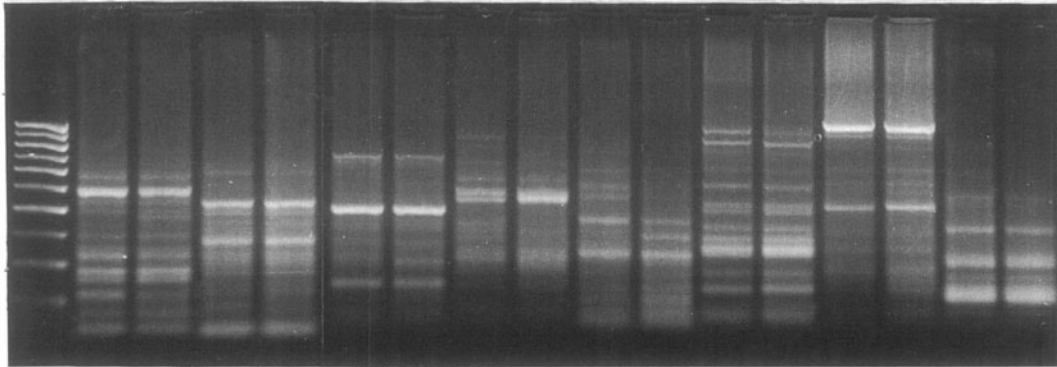


Figure 5. STS profiles of two bread wheat genotypes (1, 2) obtained using eight different STS primers (two adjacent lanes for each primer).

PCR amplified products obtained through the use of STS or SCAR primers. The technique is then described as Cleaved Amplified Polymorphic Sequences (CAPS).

2.2.2.2 Direct Amplification of Length Polymorphism (DALP)

This recently developed marker system is designed to detect polymorphic loci with co-dominant alleles, and involves following three steps: (i) PCR amplification using a pair of arbitrary primers (~20 nucleotides long) and size separation of amplified products by PAGE; (ii) isolation, re-amplification and direct sequencing of individual bands, and (iii) designing of locus specific primers for PCR analysis (Desmarais *et al.*, 1998). DALP markers developed thus have been successfully used to demonstrate elimination of paternal genome in male tissues of a mite and were found to be co-dominant (Perrot-Minnot *et al.*, 2000). This marker system is comparable to AP-PCR/RAPD markers converted to SCARs except that a pair of primers, each ~20 bases long, is used instead of a single primer that is 20 bases long in AP-PCR and 10 bases long in RAPD. DALP marker system has not been used so far in any plant material, but may be useful in future for developing locus specific polyallelic, co-dominant markers, like SSRs without the need for locating sequences containing SSRs.

2.2.2.3 Simple Sequence Repeats (SSRs)

SSRs, also known as Short Tandem Repeats (STRs) or microsatellites (1-6 bases long), are ubiquitous in eukaryotic genomes and can be analysed through PCR technology. The sequences flanking specific microsatellite loci

in a genome are believed to be conserved within a particular species, across species within a genus and rarely even across related genera. These flanking sequences, therefore, have been used to design primers for individual microsatellite loci (for details see later) and the technique is described as Sequence Tagged Microsatellite Site (STMS) analysis (Beckmann and Soller, 1990) or as Simple Sequence Length Polymorphism (SSLP) (Tautz, 1989).

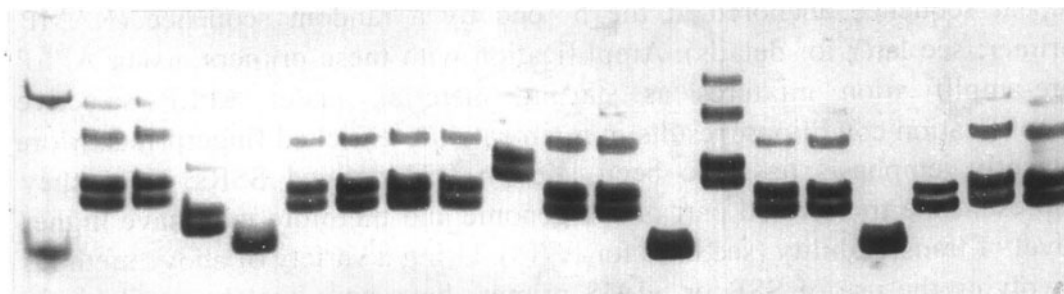


Figure 6. Amplification profiles in 19 wheat genotypes obtained using one wheat microsatellite primer pair (reproduced from Prasad *et al.*, 2000, with permission).

The STMS or SSR markers reveal polymorphisms due to variation in the lengths of microsatellites at specific individual loci; they are therefore, polyallelic and co-dominant in nature, thus proving to be very useful (Fig. 6).

Consequently, they have been used extensively not only for mapping SSR loci in many crop plants, but also for tagging a number of genes in these crops (for details, consult Gupta and Varshney, 2000). Recently, SSRs have also been used to anchor BAC clones thus facilitating their physical mapping in cotton (Cantrell *et al.*, 2001) and barley (L. Altschmied, IPK- Gatersleben, personal communication).

Since development of SSR markers requires cloning and sequencing, initially it is very costly and labour-intensive, but once the locus specific primers become available, the approach becomes cost-effective. STMS primers, based on sequences flanking individual microsatellites, can be developed either through a search for microsatellites in the DNA sequence databases (including genomic sequences and cDNA sequences) or through sequencing of restriction fragments or clones carrying microsatellites. In the latter case, the designing of primers from sequence of a clone sometimes becomes problematic, when an SSR is found to be too close to an end of the sequenced clone, leaving no room for designing primers on the basis of

sequences flanking the microsatellite. In such a situation, an anchored primer can be designed using SSR with its associated sequence and the second specific primer can be designed from a sequence, 200-300 nucleotides away from the SSR (Varghese *et al.*, 2000). A similar modified form of STMS involving AFLP (see later for details) and known as M-AFLP (microsatellite-AFLP) has also been used to generate STMS markers in *Poa* (Albertini *et al.*, 2001). This M-AFLP approach is based on the use of an AFLP primer (*Eco*RI+ 3 bases) in combination with a primer consisting of microsatellite repeat sequence anchored at the 5' end by a random sequence (RAMP primer; see later for details). Amplification with these primers, using AFLP pre-amplification mixtures as starting material, under AFLP selective amplification conditions, results in microsatellite-enriched fingerprints. More recently, emphasis has also been laid on EST-derived SSRs, since they represent the transcribed part of the genome and therefore may have higher level of transferability (see later for ESTs). Using a variety of above methods involving the use of SSR or STMS primers have now become available in several crops including rice and bread wheat. The details of these methods, the crop-wise availability of STMS primers and examples of their effective use are available elsewhere (Gupta and Varshney, 2000; Maguire, 2001).

In SSR analysis, high resolution, even without applying radioactivity, can be achieved through the use of polyacrylamide gels in combination with either ethidium bromide staining (Scrimshaw, 1992) or silver staining (Klinkicht and Tautz, 1992). Both denaturing and non-denaturing PAGE were used to resolve size differences between alleles (Lagoda *et al.*, 1998; Gay *et al.*, 1999; Brondani *et al.*, 1999). The time needed for SSR analysis may be reduced by using one or more of the following approaches: (1) high throughput approaches of DNA extraction (since high quality DNA is not needed for PCR), (2) multiplexing facilitated either by size differences in the amplified products or through the use of fluorescent primers, (3) multiple loading in a series of tiers on the same slab gel, (4) laser detection through the use of automated DNA sequencers (Ziegle *et al.*, 1992; Schwengel *et al.*, 1994; Diwan and Cregan, 1997), and (5) use of capillary electrophoresis (CE) in place of conventional slab gel electrophoresis (Mitchelson *et al.*, 1997). The fluorescent dyes often used for labelling the primers to facilitate laser detection include the following: (i) 6-carboxy-fluoresceine (FAM), (ii) hexachloro-6-carboxy-fluoresceine (HEX) and (iii) 6-carboxy-X-rhodamine (ROX) or tetrachloro-6-carboxy-fluoresceine (TET). However, the use of these fluorescent dyes makes the high throughput approach rather expensive, since labelling of one primer in the 50 nmol range may cost as much as US \$ 100-130. Therefore, an alternative high throughput approach for resource

poor laboratories, has recently been suggested (Schuelke, 2000), where for a two step PCR amplification, following three primers have been successfully employed: (i) a sequence specific forward primer, with M13 (-21) tail at its 5' end, (ii) a sequence specific reverse primer and (iii) a universal FAM labelled M13 (-21) primer. In the first few cycles of PCR amplification, forward primer with M13 (-21) is incorporated in the amplified product, which facilitates the use of FAM labelled universal primer in subsequent cycles. The final amplified product can be detected using laser detection system. In this approach, the use of a universal primer with a fluorescent tag substantially reduces the cost.

2.2.2.4 Intron Fragment Length Polymorphism (IFLP)

Intron sequences in specific genes have also been used for designing primers that can be used as markers to detect differences in the lengths of specific introns in these genes. This attribute has been used to develop markers for SAD (stearoyl-acyl carrier protein desaturase) genes in sunflower (*Helianthus annuus* L.), and has been described as DNA Fragment Length Polymorphism or DFLP (Hongtrakul *et al.*, 1998).

2.2.2.5 Expressed Sequence Tags (ESTs)

An EST is a DNA sequence from a cDNA clone that corresponds to a mRNA or a part thereof. It has been shown that ESTs, 150 to 400 bp in length, are useful for search of similarity and for genome mapping. High throughput approaches have also been used for producing ESTs. One such method is SAGE (Serial Analysis of Gene Expression), which generates a number of ESTs (10-14 base pairs long) through the same sequencing reaction. Utilizing these opportunities, ESTs are being generated in several mammals and in a variety of plant systems. In humans, cDNA libraries derived from mRNA have been utilized for generation of as many as 3,844,079 ESTs (dbEST, Oct. 12, 2001). In plants also, a large number of ESTs have been generated (Table 2) and some of them have also been mapped. For instance, in wheat, through the efforts of an International EST Triticeae Co-operative (IETC), more than 60,000 ESTs have already become available in the public domain (another more than 100,000 ESTs in wheat are available with the private sector), and a target of as many as 300,000 ESTs has been fixed each for wheat and barley (ITMI meeting, June 14-16, 2000; University of Delaware, USA). Several companies also possess large

Table 2. A summary of number of ESTs belonging to important plant genomes available in dbEST database (release, October 12 , 2001)

Plant species (common name)	Number
<i>Glycine max</i> (soybean)	195, 409
<i>Medicago truncatula</i> (barrel medic)	137, 588
<i>Lycopersicon esculentum</i> (tomato)	126, 756
<i>Arabidopsis thaliana</i> (thale cress)	113, 330
<i>Zea mays</i> (maize)	106, 595
<i>Oryza sativa</i> (rice)	92, 368
<i>Sorghum bicolor</i> (sorghum)	76, 645
<i>Hordeum vulgare</i> (barley)	73, 480
<i>Triticum aestivum</i> (wheat)	68, 379
<i>Solanum tuberosum</i> (potato)	53, 876
<i>Gossypium arboreum</i> (asiatic cotton)	20, 978
<i>Gossypium hirsutum</i> (upland cotton)	9, 438
<i>Secale cereale</i> (rye)	8, 123
<i>Brassica napus</i> (oilseed rape)	1, 774
<i>Brassica campestris</i> (field mustard)	963

private EST databases for various crop plants (e.g. wheat, maize and soybean), the access to which can be negotiated on a case-by-case basis.

In any organism, EST databases can also be used for the preparation of cDNA microarrays for a study of the functions of all specific gene sequences and their expression in time and space. These microarrays can be used for hybridization leading to the preparation of expression profiles of genes in different organs of a plant system (Ruan *et al.*, 1998). Such a study has recently been conducted in barley where a set of ESTs was selected to identify the genes specifically expressed in maternal and filial tissues of barley caryopses (Sreenivasulu *et al.*, 2001a). Using nucleotide sequence database (GenBank) and protein sequence database (PIR), ESTs can also be matched with existing sequences and assigned to specific genes with the help of available softwares. It has been shown that more than one half of the total set of *Arabidopsis* genes are likely to be represented in EST database (Bouchez and Hofte, 1998), and a massive project described as “The 2010 Project” has been initiated in USA, to decipher the functions of all 25,000 genes in this genome. The large-scale cDNA analysis is in progress in many triticeae species like rice (Yamamoto and Sasaki, 1997; Ewing *et al.*, 1999), barley (Michalek *et al.*, 2001), wheat (Ogihara *et al.*, 2001) and maize (Gallavotti *et al.*, 2001) with an aim to catalogue all expressed genes in each of these cereals including genes, which have tissue-specific, developmental

stage specific, and stress specific expression. They have also provided markers for genetic/physical mapping in plant species like soybean (Matthews *et al.*, 2001) and barley (Wolf *et al.*, 2001). Although, in an earlier study involving analysis of colinearity between *Arabidopsis* and cereal genomes has shown that colinearity with genomes like that of *Arabidopsis* may not always be sufficient to allow cross-species gene prediction and isolation in rice and other cereals (Devos *et al.*, 2000), EST databases and EST clones did prove useful in some cases for interspecies gene discovery and sequence comparison. For instance in a pilot study, a set of ESTs, generated in barley, has been successfully used to identify genes in foxtail millet (Sreenivasulu *et al.*, 2001b).

2.2.2.6 EST-derived STSs, SSRs and SNPs

Due to current emphasis on functional genomics, ESTs in large number are fast accumulating in EST databases (dbEST). They can be converted into STSs and tested for polymorphism, so that polymorphic STS markers can be developed from EST databases (Gilpin *et al.*, 1997). SSRs and SNPs can also searched in these EST databases and used for designing locus specific primers. In the past, development of SSR SNP markers has been expensive, but now EST derived SSRs and SNPs are free by-product of the currently expanding EST databases. These SSRs and SNPs are obviously limited to those species for which sufficient number of ESTs from a number of genotypes of a crop exists in the database. To date this specific approach has been used for rice (Miyao *et al.*, 1996; Cho *et al.*, 2000) and grapes (Scott *et al.*, 2000). However, generation and mapping of such EST-derived SSRs and SNPs (see later) is currently in progress in some important crop species like wheat (Eujayl *et al.*, 2000; Bundock *et al.*, 2001; Somers *et al.*, 2001) and barley (Henry *et al.*, 2000; Bundock *et al.*, 2001; Wolf *et al.*, 2001). EST-derived SSRs have some intrinsic advantages over genomic SSRs, since they are quickly obtained by electronic sorting, are unbiased in repeat type, are present in gene rich regions of the genome, and are still abundant (Scott, 2001). The usefulness of these EST-derived SSRs also lies in their expected transferability, since they are based on the conserved coding region of the genome.

2.2.2.7 Random Amplified Microsatellite Polymorphism (RAMP) and Retroposon-Microsatellite Amplified Polymorphism (REMAP)

In this approach, amplification is performed using a pair of primers, of which one primer is an anchored microsatellite (SSR) and the other primer can be

either a RAPD primer (as in RAMP) or a retroposon LTR primer (as in REMAP; Fig. 4b). In a study in barley, while using REMAP, *Copia*-specific BARE-1 primer (BARE-1 retroposon was isolated from barley) was used with an anchored SSR primer, so that the acronym *Copia*-SSR was used for this technique (Provan *et al.*, 1999). In both these approaches (RAMP and REMAP), the amplified products resolve length polymorphism that may be present either at the SSR target site itself, or at the associated sequence between the binding sites of the two primers. The RAPD/retroposon primer binding site actually serves as an arbitrary endpoint for the SSR-based amplification product, and therefore the products obtained are not as restricted by the relative genomic positions of a specific SSR as they are with ISA/ ISSR, where a single anchored primer is used. The amplified products may also be digested with a restriction enzyme, to resolve further the polymorphism for genetic mapping, as done in barley (Becker and Heun, 1995), so that it is then described as ‘digested RAMPs (dRAMPs)’. The merit of RAMP and REMAP lies in the fact that undigested total genomic DNA is used as a template, instead of preamplified restriction digested DNA used in AFLP/SAMPL/S-SAP (to be discussed in the next section)

2.2.2.8 Amplified Fragment Length Polymorphism (AFLP) and its Modified Forms

AFLP is based on PCR amplification of a set of restriction fragments, selected from a pool of fragments, that are generated due to digestion with a pair of specific restriction enzymes, one of them being a frequent cutter (e.g. *MseI*), and the other being a rare cutter (e.g., *EcoRI*). To facilitate designing of primers for selection of restriction fragments, oligonucleotide adapters, few base pairs (~20) long are ligated at the ends of these DNA fragments (Vos *et al.*, 1995). With the help of these ligated adapters, the number of DNA fragments to be amplified can be restricted, since the primers are designed to bind to the following sequence: ligated adapters + the restriction site of the enzyme used for digestion + 1-3 selective bases (chosen randomly). This method generates a large number of bands (but not too many to become difficult to score), representing the amplified products from selected restriction fragments (Fig. 7a), thus facilitating the detection of polymorphism (for details, see Vos *et al.*, 1995). It has been shown that in AFLP, a single primer combination detects up to 8 times more polymorphic bands and up to 16 times more loci, when compared with RFLP (Mackill *et al.*, 1996; Maughan *et al.*, 1996). A comparison of different molecular marker techniques (RFLP, RAPD, SSR and AFLP) also demonstrated that

AFLP is the most efficient technique for detecting polymorphism (Powell *et al.*, 1996; Linn *et al.*, 1996; Ma and Lapitan, 1998). Consequently, AFLP markers have been used for a variety of purposes including the preparation of molecular maps (Becker *et al.*, 1995; van Eck *et al.*, 1995; Keim *et al.*, 1997; Maheswari *et al.*, 1997; Castiglioni *et al.*, 1998; Qi *et al.*, 1998; Zhu *et al.*, 1998; Mank *et al.*, 1999), assessing genetic diversity (Hill *et al.*, 1996; Powell *et al.*, 1996; Milbourne *et al.*, 1997; Paul *et al.*, 1997; Zhu *et al.*, 1998; Barrett and Kidwell, 1998; Xu *et al.*, 2000; Larson *et al.*, 2001), gene tagging (Brahm *et al.*, 2000) and also for association genetics (Pakhinyat *et al.*, 1997). Many diagnostic molecular markers for different traits have also been identified through AFLPs in different crops including bread wheat (see Gupta *et al.*, 1999b). The polymorphic bands from AFLP patterns may also be converted into RFLP probes or locus specific PCR markers (SCARs), although problems may be encountered in the latter (Shan *et al.*, 1999).

AFLP generally makes use of genomic DNA, but cDNA can also be used, so that AFLP is then described as cDNA-AFLP, which has been used for transcript profiling/differential gene expression studies (Bachem *et al.*, 1996; Liscum *et al.*, 1997). The major advantages of the use of cDNA for AFLP are: (a) no prior gene or EST sequence information is required, thus making it a particularly useful gene discovery tool, (b) a large fraction of expressed genes is targeted, (c) the technique is very sensitive and is able to detect low abundance transcripts, and (d) transcript fragments can be easily extracted from gels and sequenced. Keeping in view the above, Bachem *et al.* (1998) showed that the cDNA-AFLP method serves as a robust and reproducible method for the routine detection of differentially expressed transcripts in a wide range of experimental systems. This technology is being used in several plant systems including wheat (Ciaffi *et al.*, 2001) and barley (Peter Langridge, personal communication).

The technique of AFLP has been variously modified, where an AFLP primer with 3 selective nucleotides is used either in combination with a microsatellite primer (Selective Amplification of Microsatellite Polymorphic Loci = SAMPL; Morgante and Vogel, 1994) or in combination with a retroposon based primer (Sequence-Specific Amplification Polymorphism = S-SAP; Waugh *et al.*, 1997). Generally, SAMPL primer (18-20 nucleotides long) is based on sequences of two different adjacent SSRs and an associated intervening sequence, known to be found in compound repeats (as shown in corn and soybean), although SAMPL primers each based on a single SSR (with or without a small non-microsatellite sequence at 5' end for anchoring)

have also been used (see Witsenboer *et al.*, 1997). In fact, many different combinations of restriction endonucleases, selective adapter primers and SSR primers are possible for SAMPL, allowing for the detection of nearly limitless number of SSR-based polymorphisms in a genome. In a crop like wheat, where AFLP produces too many crowded bands to score, SAMPL gives relatively fewer bands, but resolves higher level of polymorphism, thus making this technique more rewarding and user friendly (Fig. 7b). Consequently, SAMPL has been used for a variety of studies, which include linkage with economic traits (Vivek and Simon, 1999), genetic diversity studies (Witsenboer *et al.*, 1997) and DNA fingerprinting (Paglia and Morgante, 1998). Recently, SAMPL has been successfully used in wheat for the first time for gene tagging, genotype identification and genetic diversity studies at our laboratory (Roy *et al.*, 2001). In a more recent study, which is the first of its kind involving use of SAMPL in domestic animals, SAMPL and AFLP have been used for determining genetic variability among sheep breeds (Bogan *et al.*, 2001). S-SAP, however, has so far been successfully used in a solitary study involving analysis of the distribution of BARE-1 like retroposons in barley genome (Waugh *et al.*, 1997). However, there is no reason, why S-SAP will not be utilized in future for gene tagging, genome mapping and for DNA fingerprinting.

In several of the above studies, it has also been shown that SAMPL (presumably also S-SAP) markers, like AFLP, are also often dominant, but since in case of SAMPL, the polymorphism can often be due to SSR length polymorphism, a fraction of SAMPL markers can also be co-dominant. Independent data (segregation, sequence information), however, may be needed to confirm the co-dominant nature of these markers (Witsenboer *et al.*, 1997). A polymorphic band from a SAMPL or S-SAP fingerprint gel can also be converted into a conventional, single locus PCR-based marker by cloning and sequencing.

2.2.3 Single Strand Conformation Polymorphism (SSCP)

Single Strand Conformation Polymorphism (SSCP) technology allows detection of polymorphism due to differences of one or more base pairs in the PCR products. The technique relies on the secondary structure being different for single strands derived from PCR products that differ by one or more nucleotides at an internal site within the strand. In order to detect such differences, PCR products are denatured and electrophoretically separated in neutral acrylamide gels (Orita *et al.*, 1989a, b). PCR products that do not

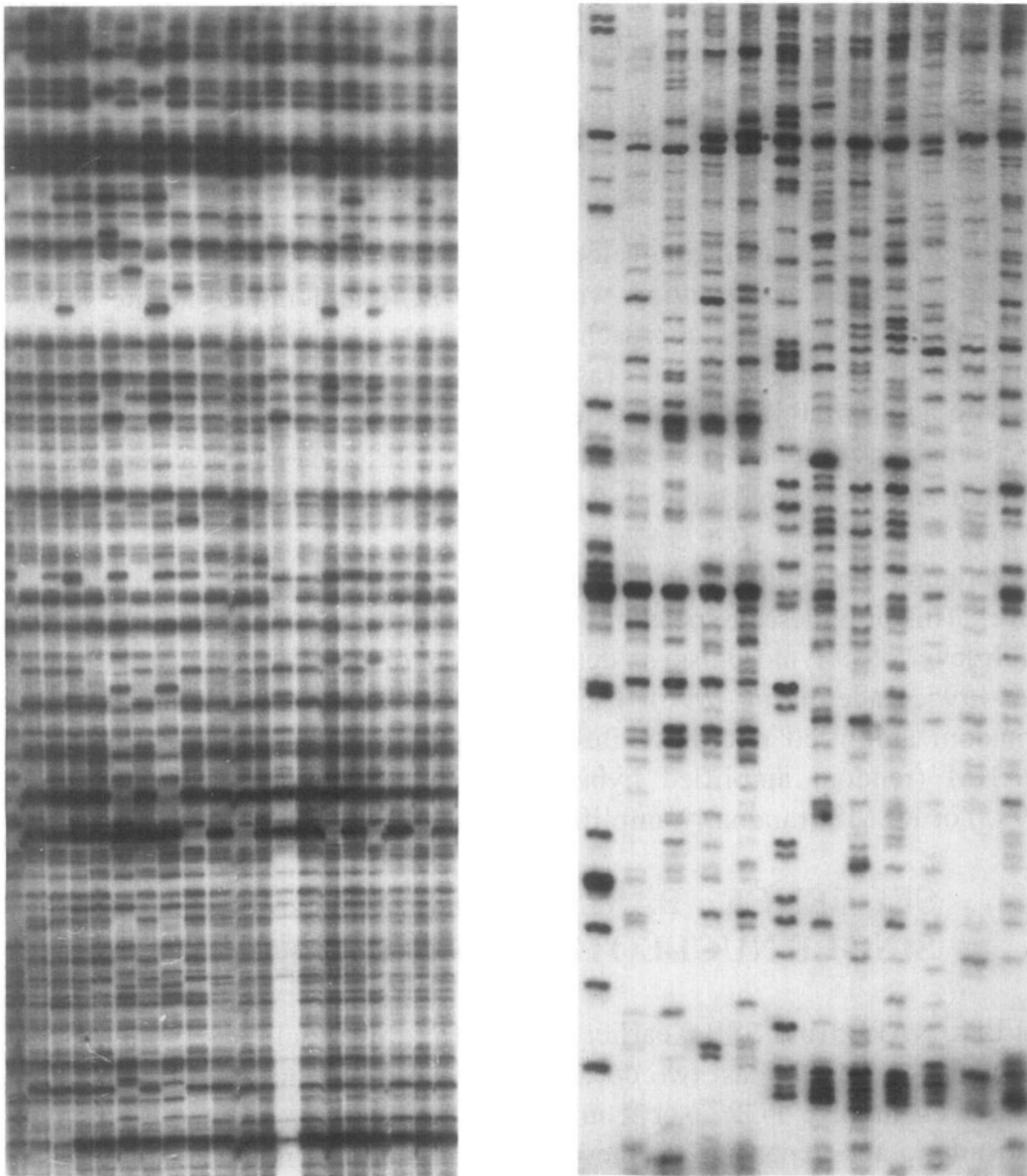


Figure 7. Amplification profiles in 20 and 12 wheat genotypes obtained using AFLP (a) and SAMPL (b) primers, respectively.

differ in fragment length have been shown to exhibit SSCP in studies conducted in sunflower (Hongtrakul *et al.*, 1998) and pine (Plomion *et al.*, 1999). Since only the PCR products are subjected to SSCP, sequence information is often needed for designing the PCR primers and isotopic labelling and autoradiography are used to detect the SSCP variants. This makes the technique relatively unsuitable for routine mapping or tagging studies, although in its sensitivity to detect polymorphism, only complete sequencing excels it.

2.3 MOLECULAR MARKERS BASED ON PCR FOLLOWED BY HYBRIDIZATION

2.3.1 Oligonucleotide Fingerprinting Using RAPD/MP-PCR Fragments as Probes

A novel strategy was also developed which combines several advantages of oligonucleotide fingerprinting with RAPD-PCR and microsatellite primed PCR (MP-PCR). In this approach, genomic DNA is amplified with either a single arbitrary 10-mer primer (as in RAPD analysis; Williams *et al.*, 1990) or with a microsatellite- complementary 15-mer or 10-mer primer (as in MP-PCR analysis; Gupta *et al.*, 1994) and the PCR products are electrophoresed, blotted and hybridized to a $\gamma^{32}\text{P}$ or digoxigenin- labelled SSR [eg. (CA)₈, (GA)₈, (GTG)₅, (GCGA)₄]. Subsequent autoradiography reveals reproducible, probe dependent fingerprints, polymorphic at the intraspecific level. This provides for speed of the assay along with high sensitivity, so that high level of polymorphism is detected (Richardson *et al.*, 1995; Cifarelli *et al.*, 1995; Ender *et al.*, 1996). The method was variously termed as RAMPO (random amplified microsatellite polymorphism; Richardson *et al.*, 1995), RAHM (random amplified hybridization microsatellites; Cifarelli *et al.*, 1995) or RAMS (randomly amplified microsatellites; Ender *et al.*, 1996).

2.4 SEQUENCE BASED MOLECULAR MARKERS: SINGLE NUCLEOTIDE POLYMORPHISMS (SNPs)

RFLPs, RAPDs and SSRs, earlier described in this article, were the markers of choice during the last two decades, but all of them need gel based assays and often involve PCR except in case of RFLP. They are, therefore, time consuming and expensive. In view of this, emphasis is now shifting towards the development of molecular markers, which can be detected through non-gel based assays, and if possible without PCR amplification. The most popular marker system, which involves detection by gel-based/non-gel based assay, sometimes even without the use of PCR amplification, is Single Nucleotide Polymorphisms (SNPs; often pronounced as 'snips'), which represent sites, where DNA sequence differs by a single base {some workers prefer to call them "simple nucleotide polymorphism" to include in this class 'indels' (insertions and deletions) and other sequence variations due to few bases}. In human genome, this polymorphism has been shown to be the most abundant with an estimated average frequency of one SNP per kilobase pairs,

so that there should be approximately 3000,000 SNPs in the entire human genome (Wang *et al.*, 1998). The development of these SNPs involving non-gel based assays has recently been also facilitated by the availability of genome-wide sequences and EST databases. For instance, SNP pipeline was recently utilized for the identification of ~3000 candidate SNPs from human EST database (Buetow *et al.*, 1999). The high density oligonucleotide arrays on DNA chips, the use of MALDI-TOF MS, and pyrosequencing that recently became available, also allow genotyping at large number of these biallelic loci in parallel, since it requires only plus/minus assay, permitting easier automation. The approach used for this purpose, relies on the capacity to distinguish a perfect match from a single base mismatch (for more details, see Gupta *et al.*, 1999a). Although, SNPs also suffer with some disadvantages, being biallelic as against polyallelic SSRs, their abundance makes them more attractive. For instance, at least in some parts of the human genome, the frequency of SNPs has been shown to be an order of magnitude higher than that of SSRs (Kwok *et al.*, 1996). There is also some evidence that the stability of SNPs and, therefore, the relative fidelity of their inheritance is higher than other marker systems like SSRs and AFLPs. These advantages led to rapid development of a number of methods for SNP detection, leading to construction of a human SNP genetic map that is being expanded. In future, similar maps will certainly be prepared and used extensively in many plant systems. A detailed account on discovery, genotyping and applications of SNPs in plants is available in a recent book (Bhatramakki and Rafalski, 2001; Edwards and Mogg, 2001; Lemieux, 2001). Few reports are also available on discovery/assessment/ frequency of SNPs in some plant systems like soybean (Grimm *et al.*, 1999), maize (Rafalski *et al.*, 1999, Ching and Rafalski *et al.*, 2000), sugar beet (Schneider *et al.*, 2001); wheat (Martin-Lopes *et al.*, 2001; Somers *et al.*, 2001; Bundock *et al.*, 2001), barley (Kota *et al.*, 1999; Henry *et al.*, 2000; Neuhaus *et al.*, 2001; Bundock *et al.*, 2001) and Norway spruce (Ivanisovich *et al.*, 2001). The genome wide mapping of SNPs has been conducted in *Arabidopsis* (Cho *et al.*, 1999) and is in progress in barley (Kota *et al.*, 2001).

2.4.1 Gel-Based Assays of PCR Products for SNP Detection

The presence of an SNP can be detected by RFLP or AFLP conducted on PCR products, whenever such an SNP generates or destroys a specific restriction site for an enzyme. This method for the detection of SNPs still needs gel based assay, and therefore, has recently been taken over by several other methods involving non-gel based assays. For instance, if SNP is

present at 3' end of an amplicon template, it can be detected simply by the failure of amplification due to mismatch between the primer sequence and the binding site in the template, although it may be difficult to distinguish this failure of PCR due to SNP from PCR failure due to other reasons. Other non-gel based approaches, that are available for the detection of SNPs at the internal position of an amplicon generally need initial PCR amplification, followed by a non-gel based assay, that discriminates between wild and mutant alleles. There are still other assays available now, which neither need separation on the gel, nor do they need PCR amplification (e.g. "invasive cleavage assay").

2.4.2 Non-Gel Based Assays of PCR Products for SNP Detection

The common non-gel based assays for detection of SNPs at the internal sites are based on the detection of mismatch between the PCR product and an oligonucleotide used as a probe. Following are some examples of such assays: (1) In an assay described as TaqMan, an oligonucleotide probe is labelled with a fluorescent reporter molecule (eg. FAM or TET) at 5' end and a quencher (e.g. TAMRA) at the 3' end. The TaqMan probe after hybridization to the template DNA is degraded at its 5' end during extension phase of PCR due to exonuclease activity of Taq polymerase enzyme (TaqMan polymerase), so that the reporter is released leading to a rise in fluorescence signal. However, when due to the presence of an SNP, the probe mismatches with the template leading to failure in duplex formation, no such degradation at 5' end of the probe is possible and there is no rise in fluorescence signal (Livak *et al.*, 1995). Different combinations of reporters and quenchers, will also permit multiplexing so that as many as six SNPs can be scored in a single PCR reaction (Lee *et al.*, 1999). (2) In another assay described as Molecular Beacon (Tyagi and Kramer, 1996; Tyagi *et al.*, 1998), the oligonucleotide probe (molecular beacon) consists of the target SNP sequence, with sequences at its two ends being complementary to each other. The two ends of the oligonucleotide are labelled just like the oligonucleotide probe used in TaqMan assay. The probe, when fails to form a duplex with the template DNA, generates a hairpin structure due to self-annealing of its two ends, thus quenching the reporter. But when the probe anneals with the template, it gets linearized, thus separating the reporter from quencher and permitting fluorescence signal. The fluorescence signals, both in TaqMan and Molecular Beacon can be detected by appropriate sensing devices. Molecular beacons have been used and recommended for

diagnostics in humans but have also been used for analysis of SNPs in crop plants like barley (Kota *et al.*, 1999). (3) In yet another approach, known as “*Oligonucleotide Ligation Assay* or *OLA*” (Landegren *et al.*, 1988), two independent probes (one is 5′ biotinylated and the other 3′ labelled) are used for hybridization with PCR product, so that, when the probes match the product, the two probes anneal with the PCR product and undergo ligation resulting in an oligonucleotide which is biotinylated at 5′ end and labelled at 3′ end. The ligation product, which is labelled at 3′ end, is captured on a solid streptavidin-coated matrix and signal is detected by autoradiography. However, when there is a mismatch due to the presence of an SNP, labelled and biotinylated oligonucleotides are unable to ligate, so that when captured by streptavidin, they carry no signal. (4) DNA chip with microarrays of oligonucleotides of known overlapping sequences, which differ at specific solitary nucleotides (at the site of SNP) can also be used for detection of SNPs. The technique is also described as sequencing by hybridization (SBH) and involves tiling strategy. Four oligonucleotides will differ at the SNP position and only one will be fully homologous. When such an array is hybridized with biotinylated PCR product, the perfect match will allow binding and mismatched products will be washed away. The perfect match can be detected through a suitable detection system (for details, see Gupta *et al.* 1999a; Sapolsky *et al.*, 1999). As many as 300,000 oligonucleotides can be available on a 1.3 x 1.3 cm² slide, permitting many SNPs to be analysed in parallel. This technique is not suitable for genotyping many individuals at one or few SNP sites, as needed for MAS, but is suitable for detecting new SNPs. (5) *Dynamic Allele-Specific Hybridization (DASH)*: This technique is based on the differences in melting temperatures between duplexes resulting due to perfect match and mismatch between the PCR product and an oligonucleotide, 15-21 base long (Howell *et al.*, 1999). The PCR product is immobilized on a solid support and denatured to give single stranded DNA, which is then hybridized to an oligonucleotide probe containing the SNP site. The duplex formed is detected by an intercalating fluorescent dye specific for double stranded DNA. When increasing temperature denatures this hybrid duplex, the melting can be followed by the reduction in the fluorescence, which depends upon the relative proportion of duplex DNA and single stranded DNAs. A rapid and sudden fall in fluorescence will indicate the T_m, and a dramatic fall in T_m will suggest a mismatch, thus identifying the SNP. (6) *Minisequencing*: SNPs can also be detected by minisequencing through Sanger’s dideoxynucleotide method, where the oligonucleotide primer has a sequence one or more base pairs upstream of

the SNP site, and all the four dNTPS along with one ddNTP corresponding to the SNP locus, are used for primer extension, so that the incorporation of a single ddNTP will allow detection of SNP. In some cases, one-step primer extension is achieved through the use of a primer, which is just upstream of the SNP site, so that incorporation of a single ddNTP will terminate the reaction and will allow the detection of SNP (Syvanen *et al.*, 1990). One such technique described as Genetic Bit Analysis (GBA) is based on hybridization capture of a single stranded PCR product to a sequence-specific microplate-bound primer, followed by enzyme-mediated single base extension of the captured

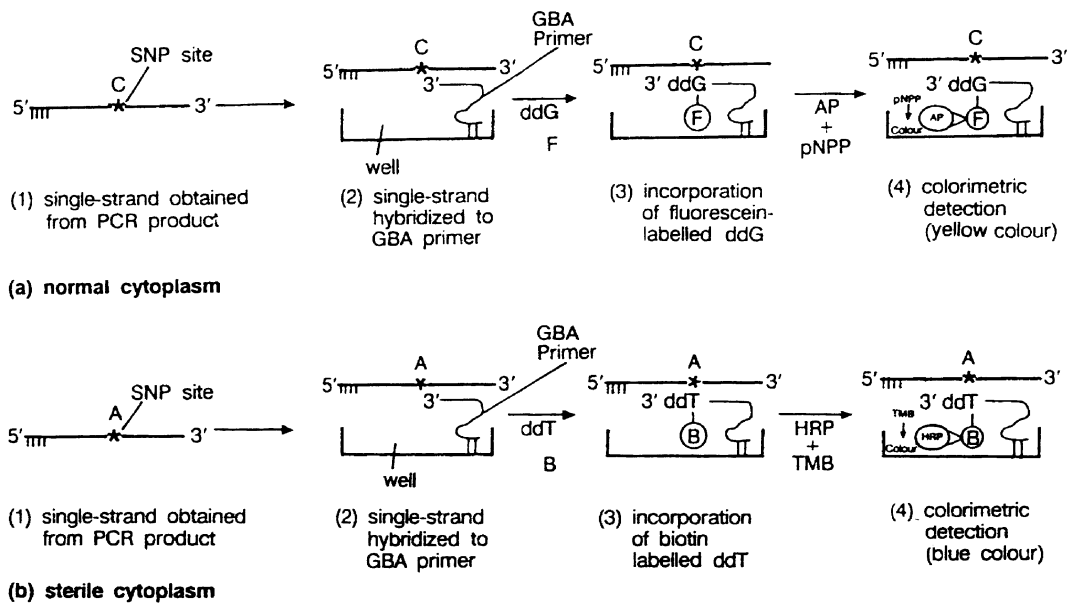


Figure 8. Schematic representation of single-nucleotide typing using GBA in onion (see text for details) (modified from Alcalá *et al.* 1997).

primer across the polymorphic site, enabling direct determination of SNP through colorimetry (Nikiforov *et al.*, 1994). The technique has been used as a diagnostic tool in human paternity tests as well as in pedigree analysis. Among plant systems, the successful use of the technique has been demonstrated in onions for distinguishing between plastomes of cytoplasmic male sterile (CMS) and fertile lines, which differ by a single SNP (Fig. 8). It has been shown that semi-automated GBA can be conducted using 96 well microtiter plates (Alcalá *et al.*, 1997). (7) *Pyrosequencing for SNP genotyping*. Pyrosequencing is a new method of obtaining short segments of sequences (upto ~20 nucleotides), simultaneously on 96 templates. Once the templates have been prepared, 96 templates can be sequenced in 15 minutes through an automated machine, being used by DuPont and Pioneer Hi-Bred

in USA. The method relies on step-wise addition of individual dNTP (with simultaneous release of pyrophosphates i.e. pp) and monitoring their template guided incorporation into the growing DNA chain via chemiluminescent detection of the formation of pyrophosphate in the incorporation reaction. Pyrosequencing is particularly suitable for SNP genotyping since genotyping of previously identified SNPs requires sequencing of only a few nucleotides (1-5 bp). Pyrosequencing is being used for SNP genotyping and also for rapid mapping of ESTs in wheat and corn (Ching and Rafalski, 2000). For this purpose, following steps are used: (i) procure sequence information generated in the SNP identification programme, (ii) design sequencing primers close to identified SNP sites; (iii) amplify SNP loci using one biotinylated and one standard primers; (iv) separate biotinylated single strands by magnetic strand separation on streptavidin-coated beads, (v) use of liquid phase pyrosequencing machine, with above primers and the bead bound template for pyrosequencing a few bases. (8) *Temperature Modulated Heteroduplex Analysis (TMHA) using dHPLC WAVETM system*. TMHA has an extraordinary sensitivity for distinguishing heteroduplexes from homoduplexes, and this principle is utilized for detection of SNPs. The mixture of wild type and variant DNA (SNP mutant) is heated and cooled again, so that the sample will then have a mixture of homo- and heteroduplexes. The heteroduplexes partially denature and can then be distinguished from corresponding homoduplexes by ion-pair reversed - phase liquid chromatography (IP-RP-HPLC) used in WAVETM system developed by Transgenomics for distinguishing heteroduplexes from homoduplexes (Fig. 9). This technique is being used to create a SNP map in barley at IPK- Gatersleben (Kota *et al.*, 2001).

2.4.3 A Non-PCR Invasive Cleavage Assay and MALDI-TOF MS for SNP Detection

In all the methods of SNP analysis described above, an initial target amplification involving PCR amplification is required. However, PCR does have significant limitations, when used in a high throughput approach, so that approaches involving simpler and more direct analysis of DNA without prior PCR amplification have been developed. One such novel non-PCR approach employs “*invasive cleavage assay*” and a MALDI-TOF-MS detection system (Lyamichev *et al.*, 1999; Griffin *et al.*, 1999). The invasive cleavage assay involves hybridization of genomic DNA with two sequence specific oligonucleotides, one termed *invader oligonucleotide*, and the other termed the *probe oligonucleotide*. The invader oligonucleotide has a sequence homology with a segment of genomic DNA upstream of the SNP site. The

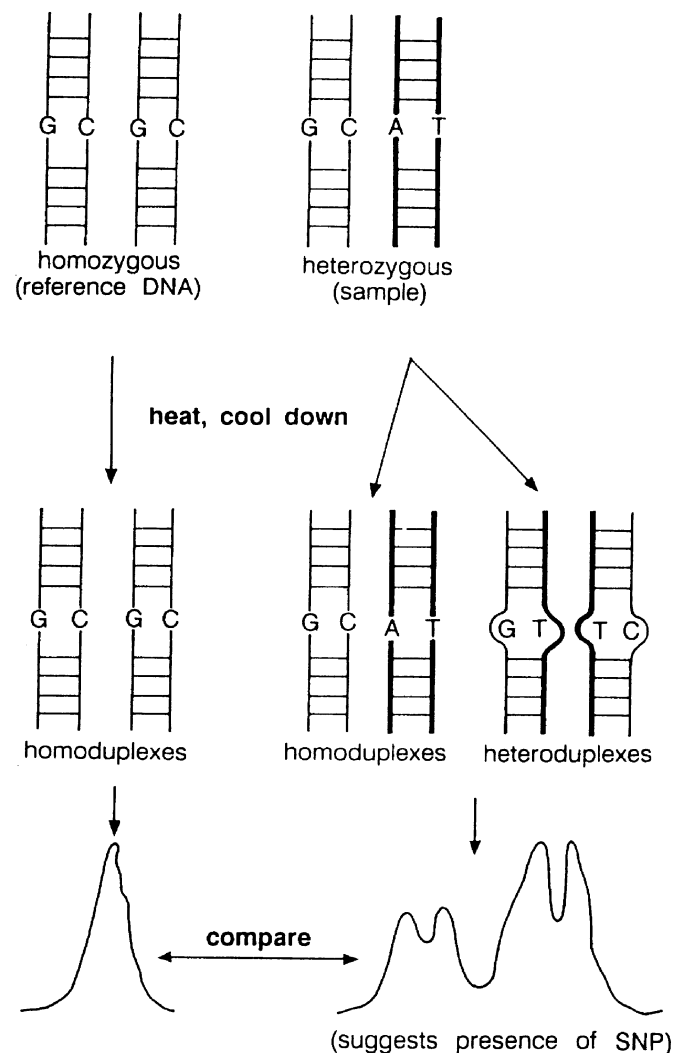


Figure 9. A diagrammatic representation of the principle involved in Temperature Modulated Heteroduplex Analysis (TMHA) using dHPLC (see text for details) (modified from Transgenomics Inc. http://arrayit.com/SNP_Services).

probe oligonucleotide, on the other hand, has a segment at its 3' end that is homologous to the target DNA, and another segment at its 5' end, that has no homology with the target DNA. On hybridization, a duplex is formed between the homologous segment of the probe oligonucleotide and the target DNA. The invader oligonucleotide now invades into the duplex for at least one nucleotide, thus forming an overlap at this point of invasion. A flap endonuclease (FEN) cleaves the unpaired region (including the overlap) on the 5' end of the probe, resulting in a 3'-hydroxyl DNA cleavage product. A modification of the above invader assay, called "*invader squared assay*" has actually been used to amplify the signal for the detection of SNPs. This is a two-steps reaction, in which primary cleavage product serves as an invader

oligonucleotide for a secondary invasive cleavage reaction, for which a fresh target and a fresh probe oligonucleotide are supplied to the reaction mix. This produces secondary cleavage products. This cycle of using cleavage product as an invader may continue for several cycles to amplify the signal, till the final product is detected by MALDI-TOF-MS (Fig. 10).

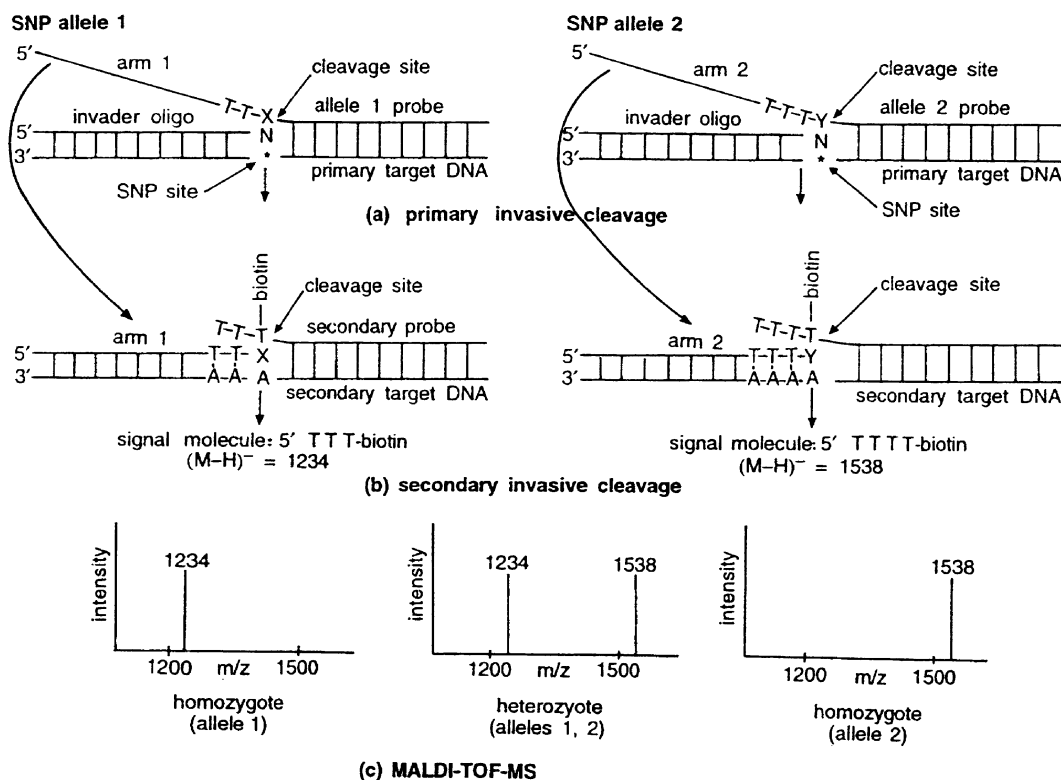


Figure 10. A diagrammatic representation of “invader squared assay” and MALDI-TOF-MS detection system for detection of SNPs (see text for details) (modified from Griffin and Smith, 2000).

3. COMPARISON OF DIFFERENT TYPES OF MOLECULAR MARKERS

In order to make a choice from a number of molecular markers that have now become available, it is necessary to make a comparison between different types of molecular markers (e.g. RFLPs, RAPDs, SSRs, AFLPs, etc.) described in this chapter. These markers have often been compared for convenience, reproducibility, speed of assay, cost effectiveness and feasibility of using automation and high throughput approaches (Table 3). However, using suitably designed experiments in several crops including soybean, barley and wheat, these markers have also been compared for their

Table 3. A comparison of general features of different types of molecular markers and their uses

	RFLPs	RAPDs	DAF	STS (CAPS)	SSRs	AFLPs	SAMPLs	REM/IRAP	SNPs
Principle	endonuclease restriction, Southern blot hybridization	DNA amplification with random primers	DNA amplification with random primers	may need endonuclease restriction of PCR products	amplification of simple sequence repeats using specific primers	endonuclease restriction use of adaptors & selective primers	same as AFLP except labelled primer is a compound SSR	DNA amplification using both/one retrotransposon & one SSR primer	sequence analysis
Types of polymorphism detected	base changes (InDel, subst.)	base changes (InDel, subst.)	base changes (InDel, subst.)	base changes (InDel, subst.)	variation in length of repeats	base changes (InDel, subst.)	base changes (InDel, subst.)	variation in length	single base changes
Sequence information needed	no	no	no	yes	yes	no	no	yes	yes
Detection method (radioisotope)	yes/ no	no	yes/ no	yes/ no	yes/ no	yes/ no	yes/ no	no	no
Reproducibility	high	low	low	high	high	high	high	high	high
No. of loci detected	1-5	1-10	20-30	1-4	1-3	>70	~50	3-5	1 (biallelic)

Inheritance	co-dominant	dominant	dominant	dominant	co-dominant	co-dominant	generally dominant	dominant / co-dominant	co-dominant	dominant
Technical difficulty	medium	low	medium	low	low	low	medium/high	medium	medium/high	dominant
Automation possible	-	-	-	+	+	+	+	+	+	+
Computation possible	+	+	+	++	++	+++	+++	+++	+++	+++
Cost	medium	low	medium	medium/high	high initially	high	high	high	low	high
Use:										
1. Varietal fingerprinting and genetic diversity	+	++	++	+	++	+++	+++	+++	++	++
2. Qualitative gene tagging	++	++	++	++	++	+++	+++	+++	+	+
3. QTL mapping	++	-	-	+	++	++	++	++	-	+
4. MAS	+	-/+	-	++	++	++	++	++	+	++
5. Comparative mapping	++	+	-	++	++	++	++	++	-	++

relative effectiveness and efficiency in detecting DNA polymorphism for a variety of purposes. For instance, utilizing two clones of poplar (*Populus × euramericana*), RAPD, AFLP and SSR were compared in eight different laboratories in Europe, each using the same genetic screening package (GSP), prepared for each marker (Jones *et al.*, 1997); it was concluded that RAPDs were not always reproducible, while AFLP and SSR gave high reproducibility. The superiority of AFLP and SSR techniques was also demonstrated in several other earlier studies (Linn *et al.*, 1996; Janssen *et al.*, 1996; Schwengel *et al.*, 1994). In another study conducted in soybean, different markers were used to estimate parameters like information content (expected heterozygosity), multiplex ratio (number of loci simultaneously analyzed per experiment) and effectiveness in assessing relationships between genotypes. A single parameter, described as marker index (product of expected heterozygosity and multiplex ratio) was also used to evaluate the relative merit of different markers. The use of this approach showed that SSR markers have the highest information content, while AFLP markers have the highest multiplex ratio and highest marker index. RAPDs were intermediate both in heterozygosity and multiplex ratio, while RFLPs had moderate heterozygosity (Powell *et al.*, 1996). Similar studies, leading to almost similar conclusions, were also conducted in barley (Russell *et al.*, 1997) and wheat (Bohn *et al.*, 1999). These experiments for comparison of different molecular markers in individual crops neither took into account the cost involved nor did they consider the feasibility of their use by practising plant breeders. Nevertheless, the comparisons are useful for laboratory experiments, where sometimes cost is not a constraint.

As mentioned above, the SSR markers have the high information content (estimated as either the polymorphic information content or as genetic diversity index or as expected heterozygosity) and are locus specific and co-dominant, thus making them the markers of choice for a variety of purposes including practical plant breeding. However, if SSR markers are developed for the first time, a very high cost of their development (due to cloning and sequencing) will restrict their use in many laboratories. But if SSR primers have already been developed in a crop, the use of SSR markers for this specific crop will be cost effective and can be used even in small laboratories (e.g., in wheat, ~1000 primer pairs are available through the efforts of WMC and individual workers). Polymorphic bands in RAPD, AFLP and SAMPL, can also be converted into locus specific SCARs, which will then prove to be as effective as SSR markers. However, for characterization of genomes

(fingerprinting) in a population, in our opinion, ISSRs, oligonucleotide fingerprinting, AFLPs and SAMPL may be more suitable, since each of them yields a large number of fragments, giving higher estimates of multiplex ratio and marker index.

4. HIGH THROUGHPUT APPROACHES IN MOLECULAR MARKER TECHNOLOGY

For genotyping a large number of entries by the most efficient molecular marker system (e.g., SSR and AFLP), high throughput approaches are absolutely necessary. This involves extraction of only a low quality DNA (suitable for PCR) in the minimum of time (using Matrix Mill or FastPrep System), followed by amplification through DNA Engine Tetrad (thermal cycler) and laser detection through fluorescent labelling and automated sequencing machines, preferably based on capillary electrophoresis. Robot operated automated pipetting stations can be another facility that would add to automation. Scale-up and automation, however, become difficult for majority of molecular markers, since most of them are gel based. Therefore, SNPs involving non-gel based detection (permitting scale-up and automation) will be the markers of choice in future. A number of available non-gel based assays for the detection of SNPs have been briefly discussed in this chapter, but only through further use of these assays in plant systems, we will know, which of these assays is cost effective and most suitable for high throughput genotyping. However, in many of these non-gel based assays, MALDI-TPF MS can be utilized for discriminating between two alternative alleles of a SNP (Griffin and Smith, 2000). Use of high-density oligonucleotide arrays is another high throughput approach, which will be extensively used in future. It has been shown that upto 1000 samples can be genotyped in one reaction using microarrays and MALDI-TOF MS. As discussed above, TMHA dHPLC WAVE™ DNA molecular platforms comprising HPLC/capillary electrophoresis system are also being used by Transgenomics (USA) for automated screening involving SNP discovery and detection.

5. ECONOMICS OF MARKER- ASSISTED SELECTION (MAS) IN PLANT BREEDING

A number of theoretical studies have been undertaken, which demonstrated relative efficiency of MAS over phenotypic selection in plant breeding (Lande and Thompson, 1990; Luo *et al.*, 1997; Moreau *et al.*, 1998; Zhang and Smith, 1992, 1993; Gimelfarb and Lande, 1994, 1995; Whittaker *et al.*, 1995, 1997; Hospital *et al.*, 1997). Lande and Thompson (1990) had also proposed selection of individual plants on the basis of an index, based on a combination of phenotype and a molecular value that is predicted with the use of a marker. There are also examples, where possibility of success through the use of MAS in actual plant breeding has been demonstrated (Young *et al.*, 1995; Young, 1999). Despite this, there is hardly any report available, where a practising plant breeder successfully used MAS leading to release of a commercial variety. In most studies advocating the benefit of using MAS in plant breeding, the additional cost associated with the marker evaluation is ignored, thus leading to overestimation of the benefit derived from MAS. Indirect selection using molecular markers actually becomes prohibitive due to several factors including the following: (i) size of segregating population, (ii) the number of replications in a trial, (iii) the number of field trials that are sometimes needed for the study of genotype \times environment interaction, and (iv) the number of QTL associated molecular markers that may need to be used simultaneously in the same population. Therefore, studies have been conducted, where cost has been taken into consideration while evaluating the utility of MAS (Xie and Xu, 1998; Moreau *et al.*, 2000). In one of these studies it was shown that relative efficiency of MAS is reduced, when cost is taken into account. Moreover, it is also possible that the strategies used in phenotypic selection may be further refined and improved in future, thus reducing further the relative benefit derived from MAS. One can only hope that further improvement and automation in marker technology will reduce the genotyping cost, thus leading to cost efficiency of MAS, which is the major use of molecular markers in crop improvement programmes. However, there is no doubt that molecular markers will certainly be put to a variety of uses in crop improvement programmes, when viewed in a wider context, even if their practical use through MAS, in the hands of plant breeders, may take some time.

6. CONCLUSION

The last two decades have witnessed a breathtaking activity in the development and use of molecular markers. Consequently, during this period, often a new and more efficient marker system has always been available, before one could finish using the latest marker system available at a particular time. This is obvious from the fact that in mammalian systems, the first and the second-generation molecular markers (RFLP, RAPD, SSR, AFLP, etc.) are becoming obsolete and there is an increasing interest in the use of ESTs and SNPs, which will be the markers of choice in future. Many of these markers will utilize capillary electrophoresis or non-gel based assays to achieve automation and high throughput. More non-PCR methods like 'invasive cleavage' will also be developed and utilized. With these developments in the field of molecular marker technology, we are also witnessing a parallel emergence of new science of genomics that includes genome-wide sequencing, bioinformatics, functional genomics and proteomics. These new research areas will certainly influence the activity involving the use of molecular markers in plant systems. Therefore, one would hope that in plant systems also, while we will keep on using the second generation molecular markers like SSRs, AFLP and SAMPL, the available methods for SNP discovery and detection will be increasingly used in future. This will facilitate the use of molecular markers in a cost-effective manner not only for plant breeding, but also for conservation, maintenance and sustainable utilization of plant genetic resources. Genome-wide sequencing and the use of bioinformatics and gene expression studies for annotation of DNA sequences available in the databases will be another area, that will keep biologists occupied during the next decade. The study of proteins, which is also gaining momentum under the banner of proteomics, will further supplement and complement these activities involving the development and use of molecular markers.

ACKNOWLEDGEMENTS

We are thankful to Council for Scientific and Industrial Research (CSIR), Government of India for providing CSIR-ES to PKG and a CSIR-SRF to RKV and to Department of Biotechnology (DBT), Government of India for financial assistance.

REFERENCES

- Akopyanz, N., Bukanov, N., Westblom, T.U. and Berg, D.E. (1992) PCR based RFLP analysis of DNA sequence diversity in the gastric pathogen *Helicobacter pylori*. *Nucl. Acids Res.* **20**, 6221-6225.
- Alcala, J., Geovannoni, J.J., Pike, L.M., and Reddy, A.S. (1997). Application of genetic bit analysis (GBATM) for allelic selection in plant breeding. *Mol. Breed.* **3**, 495-502.
- Albertine, E., Bertoli, F., Marconi, G., and Falcinelli, M. (2001) Isolation of polymorphic microsatellites in *P. paratensis* L. by using the new microsatellite- AFLP (M-AFLP) procedure, XLV Italian society of agricultural genetics- SIGA annual congress: Abstract P5.45, 26th- 29th Sept. 2001, Salsomaggiore Terme (Italy).
- Arens, P., Odinet, P., Heusden, A.W.- van, Lindhout, P., and Vosman, B. (1995) GATA- and GACA- repeats are not evenly distributed throughout the tomato genome. *Genome* **38**, 84-90.
- Bachem, C.W.B., Oomen R.J.F.J., and Visser, R.G.F. (1998) Transcript imaging with cDNA-AFLP: a step-by-step protocol. *Plant Mol. Biol. Rep.* **16**, 157-173.
- Bachem, C.W.B., van der Hoeven, R.S., de Bruijn, S.M., Vreugdenhil, D., Zabeau, M., and Visser, R.G.F. (1996) Visualization of differential gene expression during potato tuber development. *The Plant J.* **9**, 745-753.
- Barrett, B.A. and Kidwell, K.K. (1998) AFLP-based genetic diversity assessment among wheat cultivars from the Pacific Northwest. *Crop Sci.* **38**, 1261-1271.
- Bhatramakki, D., and Rafalski, A. (2001) Discovery and application of single nucleotide polymorphism markers in plants, in R.J. Henry (ed.), *Plant Genotyping: The DNA Fingerprinting of Plants*, CABI Publishing, Oxon, U. K. pp. 179-192.
- Bebeli, P.J., Zhou, Z., Somers, D.J. and Gustafson, J.P. (1997) PCR primed with minisatellite core sequences yields DNA fingerprinting probes in wheat. *Theor. Appl. Genet.* **95**, 276-283.
- Becker, J. and Heun, M. (1995) Mapping of digested and undigested random amplified microsatellite polymorphisms in barley. *Genome* **38**, 991-998.
- Becker, J., Vos, P., Kupier, M., Salamini, F., and Heun, M. (1995) Combined mapping of AFLP and RFLP markers in barley. *Mol. Gen. Genet.* **249**, 65-73.
- Beckmann, J.S., and Soller, M. (1990) Towards a unified approach to genetic mapping of eukaryotes based on sequence tagged microsatellite sites. *Bio/Technology* **8**, 930-932.
- Bogani, D., Capomaccio, S., Cappelli, K., and Sarti, F.M. (2001) Use of AFLP and SAMPL markers for the analysis of the genetic variability of three sheep groups belonging to the Appenninica, Massese and Suffolk breeds, XLV Italian society of agricultural genetics- SIGA annual congress, 26th- 29th Sept. 2001, Salsomaggiore Terme (Italy).
- Bohn, M., Utz, H.F. and Melchinger, A.E. (1999) Genetic similarities among winter wheat cultivars determined on the basis of RFLPs, AFLPs, and SSRs and their use for predicting progeny variance. *Crop Sci.* **39**, 228-237.

- Botstein, D., White, R.L., Skolnick, M., and Davis, R.W. (1980) Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *Am. J. Hum. Genet.* **32**, 314-331.
- Bouchez, D., and Hofte H. (1998) Functional genomics in plants. *Plant Physiol.* **118**, 725-732.
- Brahm, L., Röcher, T., and Freidt, W. (2000) AFLP markers for characterizing the Azuki bean complex. *Crop Sci.* **40**, 676-682.
- Brondani, C., Vianello, P., Rosana, B. and Ferreira, M.E. (1999) Conservation of SSR loci in wild and cultivated *Oryza* species, International plant and animal genome VII conference: Abstract P452, 17th -21st Jan., 1999, San Diego, CA.
- Buetow, K.H., Edmonson, M.N., and Cassidy, A.B. (1999) Reliable identification of large numbers of candidate SNPs from public EST data. *Nature Genetics* **21**, 323-325.
- Bundock, P., Holton, T.A., Christopher, J.T., and Henry, R.J. (2001) Detection of SNPs in ESTs from wheat and barley, XVIth Eucarpia congress, Plant breeding: sustaining the future: Abstract P1.13, 10th –14th Sept. 2001, Edinburgh (UK).
- Caetano-Annoles, G. (1994) MAAP- a versatile and universal tool for genome analysis. *Plant Mol. Biol.* **25**, 1011-1026.
- Caetano-Anolles, G. (2001) Plant genotyping using arbitrarily amplified DNA, in R.J. Henry (ed.), *Plant Genotyping: The DNA Fingerprinting of Plants*, CABI Publishing, Oxon, U. K. pp. 29-46.
- Caetano-Annoles, G., Bassam, B.J., and Gresshoff, P.M. (1991) DNA amplification fingerprinting: a strategy for genome analysis. *Plant Mol. Biol. Rep.* **9**, 294-307.
- Caetano-Annoles, G., and Gresshoff, P.M. (1994) DNA amplification fingerprinting using very short arbitrary mini-hairpin oligonucleotide primers. *Bio/Technology* **12**, 619-623.
- Cantrell, R.G., Lei, E., and Wing, R.A. (2001) DNA microsatellite localization to cotton BAC library, XVIth Eucarpia congress, Plant breeding: sustaining the future: Abstract P1.12, 10th –14th Sept. 2001, Edinburgh (UK).
- Castiglioni, P., Pozzi, C., Heun, M., Terzi, V., Muller, K.J., Rohde, W., and Salamini, F. (1998) An AFLP-based procedure for the efficient mapping of mutations and DNA probes in barley. *Genetics* **149**, 2039-2056.
- Ching, A., and Rafalski, A. J. (2000) Rapid EST mapping and genotyping of SNP loci using pyrosequencing, International plant and animal genome VII conference, 17th- 21st Jan., 1999, San Diego, CA (USA), p. 112 (<http://www.intl-pag.org/pag/8/abstracts/pag8419.html>).
- Cho, Y.G., Ishii, T., Temnykh, S., Chen, X., Lipovich, L., McCouch, S.R., Park, W.D., Ayres, N., and Cartinhour, S. (2000) Diversity of microsatellites derived from genomic libraries and GenBank sequences in rice (*Oryza sativa* L.). *Theor. Appl. Genet.* **100**, 713-722.
- Cho, R.J., Mindrinos, M., Richards, D.R., Sapolsky, R.J., Anderson, M., Drenkard, E., Dewdney, J., Reuber, T.L., Stammers, M., Federspiel, N., Theologis, A., Yang, W.H., Hubbell, E., Au, M., Chung, E.Y., Lashkari, D., Lemieux, B., Dean, C., Lipshutz, R.J., Ausubel, F.M., Davis, R.W., and Oefner, P.J. (1999) Genome-wide mapping with biallelic markers in *Arabidopsis thaliana*. *Nature Genetics* **23**, 203-207.
- Ciaffi, M., Paolacci, A.R., Marabottini R., and Tanzarella, O.A. (2001) Identification and characterisation of gene sequences expressed during wheat heading, XLV Italian society of agricultural genetics- SIGA annual congress, 26th- 29th Sept. 2001, Salsomaggiore Terme (Italy).

- Cifarelli, R.A., Gallitelli, M., and Celliini, F. (1995) Random amplified hybridization microsatellites (RAHM): isolation of a new class of microsatellite- containing DNA clones. *Nucl. Acids Res.* **23**, 3802-3803.
- Devos, K.M., Pittaway, T.S., Reynolds, A., and Gale, M.D. (2000) Comparative mapping reveals a complex relationship between the pearl millet genome and those of foxtail millet and rice. *Theor. Appl. Genet.* **100**, 190-198.
- Desmarais, E., Lanneluc, I., and Lagnel, J. (1998) Direct amplification of length polymorphism (DALP), or how to get and characterize new genetic markers in many species. *Nucl. Acids Res.* **26**, 1458-1465.
- Diwan, N., and Cregan, P.B. (1997) Automated sizing of fluorescent-labeled simple sequence repeat (SSR) markers to assay genetic variation in soybean. *Theor. Appl. Genet.* **95**, 723-733.
- Edwards, A., Civetello, H., Hammond, H.A., and Caskey, C.T. (1991) DNA typing and genetic mapping with trimeric and tetrameric repeats. *Am. J. Hum. Genet.* **49**, 746-756.
- Edwards, K.J. and Mogg, R. (2001) Plant genotyping by analysis of single nucleotide polymorphisms, in R.J. Henry (ed.), *Plant Genotyping: The DNA Fingerprinting of Plants*, CABI Publishing, Oxon, U. K. pp. 1-14.
- Ender, A., Schwenk, K., Stadler, T., Streit, B., and Schierwater, B. (1996) RAPD identification of microsatellites in *Daphnia*. *Mol. Ecol.* **5**, 437-441.
- Eujayl, I., Baum, M., Sorrells, M., Wolters, P., and Powell, W. (2000) Use of EST- derived microsatellites to estimate genetic diversity in commercial durum wheat, Tenth international public workshop, 14-16 June, 2000, University of Delaware, Newark, Delaware (USA).
- Ewing, R.M., Kahla, A.B., Poirot, O., Lopez, F., Audic, S., and Claverie, J.-M. (1999) Large-scale statistical analysis of rice ESTs reveal correlated patterns of gene expression. *Genome Res.* **9**, 950-959.
- Flavell, A.J., Knox, M.R., Pearce, S.R., and Ellis, T.H.N. (1998). Retrotransposon-based insertion polymorphisms (RBIP) for high throughput marker analysis. *Plant J.* **16**, 643-650.
- Gallavotti, A., Gianfranceschi L., and Sari-Gorla M (2001) Expression profiling during ear development in maize, XLV Italian society of agricultural genetics- SIGA annual congress: Abstract P1.9, 26th- 29th Sept. 2001, Salsomaggiore Terme (Italy).
- Gale, M.D., Atkinson, M.D., Chinoy, C.N., Harcourt, J., Jia, Q., Li, Y., and Devos, K.M. (1995) Genetic maps of hexaploid wheat, in S. Chen (ed.), *Proc. 8th intern. wheat genet. symp.* China Agricultural Sciencetech Press, Beijing (China), pp. 29-40.
- Gay, C., Rodier-Goud, M., Kaye, C., Pieretti, I., Billotte, N., Lebrun, P., Risterucci, A.-M., Seguin, M., D'Hont, A., and Glaszmann, J.C. (1999) Development of microsatellite enriched libraries in several tropical species, International plant and animal genome VII conference: Abstract P465, 17th -21st Jan., 1999, San Diego, CA.
- Gilpin, B.J., McCallum, J.A., Frew, T.J., and Timmerman-Vaughan, G.M. (1997) A linkage map of the pea (*Pisum sativum* L.) genome containing cloned sequences of known functions and expressed sequence tags (ESTs). *Theor. Appl. Genet.* **95**, 1289-1299.
- Gimelfarb, A., and Lande, R. (1994) Simulation of marker- assisted selection in hybrid populations. *Genetical Res.* **63**, 39-47.
- Gimelfarb, A., and Lande, R. (1995) Marker-assisted selection and marker-QTL associations in hybrid populations. *Theor. Appl. Genet.* **91**, 522-528.

- Griffin, T.J., Hall, J.G., Prudent, J.R., and Smith, L.M., (1999). Direct genetic analysis by matrix-assisted laser desorption ionization mass spectrometry. *Proc. Nat. Acad. Sci. USA*, **96**, 6301-6306.
- Griffin, T.J., and Smith, L.M. (2000) Single-nucleotide polymorphism analysis by MALD-TOF mass spectrometry. *Trends Biotech.* **18**, 77-84.
- Grimm, D.A., Denesh, D., Mudge, J., Young, N.D., and Cregan, P.B. (1999) Assessment of single nucleotide polymorphisms (SNPs) in soybean, International plant and animal genome VII conference: Abstract P243, 17th- 21st Jan., 1999, San Diego, C.A.
- Gupta, P.K., Balyan, H.S., Sharma, P.C., and Ramesh, B. (1996) Microsatellites in plants: a new class of molecular markers. *Curr. Sci.* **70**, 45-54.
- Gupta, M., Chyi, J., Romero-Severson, J., and Owen, J.L. (1994) Amplification of DNA markers from evolutionary diverse genomes using single primers of simple- sequence repeats. *Theor. Appl. Genet.* **89**, 998-1006.
- Gupta, P.K., Roy, J.K., and Prasad, M. (1999a) DNA chips, microarrays and genomics. *Curr. Sci.* **77**, 875-884.
- Gupta, P.K., and Varshney, R.K. (2000) The development and use of microsatellite markers for genetic analysis and plant breeding with emphasis on bread wheat. *Euphytica* **113**, 163-185.
- Gupta, P.K., Varshney, R.K., Sharma, P.C., and Ramesh, B. (1999b) Molecular markers and their application in wheat breeding. *Plant Breeding* **118**, 369-390.
- Hayashi (1992) PCR-SSCP: A method for detection of mutations. *Genetic Analysis: Techniques and Applications* **9**, 73-79.
- Hearne, C.M., Ghosh, S., and Todd, J.A. (1992) Microsatellites for linkage analysis of genetic traits. *Trends Genet.* **8**, 288-294.
- Heath, D.D., Iwama, G.K., and Devlin, R.H. (1993) PCR primed with VNTR core sequences yields species specific patterns and hypervariable probes. *Nucl. Acids Res.* **21**, 5782-5785.
- Henry, R.J., Holton, T.A., Kota, R., Muirhead, A., McClure, L., and Ablett, G. (2000) SSR and SNP markers derived from barley ESTs, Tenth international public workshop, 14-16 June, 2000, University of Delaware, Newark, Delaware (USA).
- Hill, M., Witsenboer, H., Zabeau, M., Vos, P., Kesseli, R., and Michelmore, R. (1996) PCR-based fingerprinting using AFLPs as a tool for studying genetic relationships in *Lactuca* spp. *Theor. Appl. Genet.* **93**, 1202-1210.
- Hongtrakul, V., Slabaugh, M.B., and Knapp, S.J. (1998) DFLP, SSCP and SSR markers for $\Delta 9$ -stearoyl-acryl carrier protein desaturases strongly expressed in developing seeds of sunflower: intron lengths are polymorphic among elite inbred lines. *Mol. Breed.* **4**, 195-200.
- Hospital, F., Moreau, L., Charcosset, A., and Gallais, A. (1997) More on the efficiency of marker-assisted selection. *Theor. Appl. Genet.* **95**, 1181-1189.
- Howell, W.M., Jobs, M., Gyllensten, U., and Brookes, A.J. (1999) Dynamic allele-specific hybridization. *Nature Biotech.* **17**, 87-88.
- Ivanissevich, S.D., and Morgante, M. (2001) Sequence diversity and SNP marker development in Norway spruce, XLV Italian society of agricultural genetics- SIGA annual congress: Abstract P1.20, 26th- 29th Sept. 2001, Salsomaggiore Terme (Italy).
- Janssen, P., Coopman, R., Huys, G., Swing, J., Bleeker, M., Vos, P., Zabeau, M., and Kersters, K. (1996) Evaluation of the DNA fingerprinting method AFLP as a new tool in bacterial taxonomy. *Microbiology* **142**, 1881-1893.

- Jones, C.J., Edwards, K.J., Castaglione, S., Winfield, M.O., Sala, F., van de Wiel, C., Bredemeizer, G., Vosman, B., Matthes, M., Daly, A., Brettschneider, R., Bettini, P., Buiatti, M., Maestri, E., Malcevski, A., Marmioli, N., Aert, R., Volckaert, G., Rueda, J., Linacero, R., Vazquez, A., and Karp, A. (1997) Reproducibility testing of RAPD, AFLP, and SSR markers in plants by a network of European laboratories. *Mol. Breed.* **3**, 381-390.
- Kalendar, R., Grob, T., Suoniemi, A. and Schulman, A.H. (1999a) IRAP and REMAP: Two new retrotransposon-based DNA fingerprinting techniques. *Theor. Appl. Genet.* **98**, 704-711.
- Kalendar, R., Nevo, E., and Schulman, A.H. (1999b) Genome diversification in wild barley (*Hordeum spontaneum*) populations revealed with retrotransposon REMAP markers. In: International plant and animal genome VII conference: Abstract P-446, pp.187. 17th -21st Jan., 1999, San Diego, CA.
- Keim, P., Schupp, J.M., Travis, S.E., Clayton, K., Zhu, T., Shi, L., Ferreira, A., and Webb, D.M. (1997) A high -density soybean genetic map based on AFLP markers. *Crop Sci.* **37**, 537-543.
- Klinkicht, M., and Tautz, D. (1992) Detection of simple sequence length polymorphisms by silver staining. *Mol. Ecol.* **1**, 133-134.
- Kota, R., Holton, T.A., and Henry, R.J. (1999) Detection of transgenes in crop plants using molecular beacon assays. *Plant Mol. Biol. Rep.* **17**, 363-370.
- Kota, R., Wolf, M., Michalek, W., and Graner, A. (2001) Application of denaturing high-performance liquid chromatography for mapping of single nucleotide polymorphisms in barley (*Hordeum vulgare* L.) *Genome* **44**, 523-528.
- Kwok, P.Y., Deng, Q., Zakeri, H., Taylor, S.L., and Nickerson, D.A. (1996) Increasing the information content of STS-based genome maps: identifying polymorphisms in mapped STSs. *Genomics* **31**, 123-126.
- Lagoda, P.J.L., Noyer, J.-L., Dambier, D., Baurens, F.-C., Grapin, A., and Lanaud, C. (1998) Sequence tagged microsatellite site (STMS) markers in the *Musaceae*. *Mol. Ecol.* **7**, 659-662.
- Lande, R., and Thompson, R. (1990) Efficiency of marker-assisted selection in the improvement of quantitative traits. *Genetics* **124**, 743-756.
- Landegren, U., Kaiser, R., Sanders, J., and Hood, L. (1988) Ligase-mediated gene detection technique. *Science* **241**, 1077-1080.
- Larson, S.R., Waldron, S.B., Monsen, S.B., John, L.St., Palazzo, A.J., McCracken, C.L., and Harrison, R.D. (2001) AFLP variation in angiospermous and dioecious bluegrasses of western North America. *Crop Sci.* **41**, 1300-1305.
- Lee, L.G., Livak, K.J., Mullah, B., Graham, R.J., Vinayak, R.S., and Woudenberg, T.M. (1999) Seven-colour, homogenous detection of six PCR products. *Biotechniques* **2**, 342-349.
- Lemieux, B. (2001) Plant genotyping based on analysis of single nucleotide polymorphisms using microarrays, in R.J. Henry (ed.), *Plant Genotyping: The DNA Fingerprinting of Plants*, CABI Publishing, Oxon, U. K. pp. 47-58.
- Linn, J.-J., Kuo, J., Ma, J., Saunders, J.A., Beard, H.S., MacDonald, M.H., Kenworthy, W., Ude, G.N., and Matthews, B.L. (1996) Identification of molecular markers in soybean: comparing RFLP, RAPD and AFLP DNA mapping techniques. *Plant Mol. Biol. Rep.* **14**, 156-169.
- Liscum, M. (1997) AFLP: nucleic acid fingerprinting with broad applications. *Mol. Biol. Newsl.* Nov. /Dec.1997: 2-3.

- Livak, K.J., Flood, S.J.A., Marmaro, J., Giusti, W., and Deetz, K., (1995). Oligonucleotides with fluorescent dyes at opposite ends provides a quenched probe system useful for detecting PCR product and nucleic acid hybridization. *PCR-Methods and Applications* **4**, 357-362.
- Luo, Z.W., Thompson, R., and Woolliams, J.A. (1997) A population genetics model of marker-assisted selection. *Genetics* **146**, 1173-1183.
- Lyamichev, V., Mast, A.L., Hall, J.G., Prudent, J.R. Kaiser, M.W., Takova, T., Sander, T.J., deArruda, M., Arco, D.A., Neri, B.P., and Brow, M.A.D. (1999) Polymorphism identification and quantitative detection of genomic DNA by invasive cleavage of oligonucleotide probes. *Nature Biotech.* **17**, 292-296.
- Ma, Z.-Q., and Lapitan, N.L.V. (1998) A comparison of amplified and restriction fragment length polymorphism in wheat. *Cereal Res. Commun.* **26**, 7-13.
- Mackill, D.J., Zhang, Z. Redona, and Colowit, P.M. (1996) Level of polymorphism and genetic mapping of AFLP markers in rice. *Genome* **39**, 969-977.
- Maguire, T.L. (2001) Producing and exploiting enriched microsatellite libraries, in R.J. Henry (ed.), *Plant Genotyping: The DNA Fingerprinting of Plants*, CABI Publishing, Oxon, U. K. pp. 29-46.
- Maheswaran, M., Subudhiu, P.K., Nandi, S.S., Xu, J.C., Parco, A., Yang, D.C., and Huang N. (1997) Polymorphism, segregation and distribution of AFLP markers in a doubled haploid rice population. *Theor. Appl. Genet.* **94**, 39-45.
- Mank, M.V.R., Antonise, R., Bastiaans, E., Senior, M.L., Stuber, C.W., Melchinger, A.E., Lubberstedt, T., Xia, X.C., Stam, P., Zabeau, M., and Kuiper, M. (1999) Two high-density AFLP linkage maps of *Zea mays* L.: analysis of distribution of AFLP markers. *Theor. Appl. Genet.* **99**, 921-935.
- Martins-Lopes, P., Zhang, H., and Koebner, R. (2001) Detection of single nucleotide mutations in wheat using single strand conformation polymorphism gels. *Plant Mol. Biol. Rep.* **19**, 159-162.
- Matthews, B.F., Devine, T.E., Weisemann, J.M., Beard, H.S., Lewers, K.S., MacDonald, M.H., Park, Y-B., Maiti, R., Lin, J-J., Kuo, J., Pedroni, M.J., Cregan, P.B., and Saunders, J.A. (2001) Incorporation of sequenced cDNA and genomic markers into the soybean genetic map. *Crop Sci.* **41**, 516-521.
- Maughan, P.J., Saghai-Maroo, M.A., Buss, G.R., and Huestis, G.M. (1996) Amplified fragment length polymorphism (AFLP) in soybean: species diversity, inheritance, and near-isogenic analysis. *Theor. Appl. Genet.* **93**, 392-401.
- Meyer, W., Michell, T.G., Freedman, E.Z., and Vilgalys, R. (1993) Hybridization probes for conventional DNA fingerprinting used as single primers in the polymerase chain reaction to distinguish strains of *Cryptococcus neoformans*. *J. Clin. Biol.* **31**, 2274-2280.
- Michalek, W., Weschke, W., Pleissner, K.-P., and Graner, A. (2001) Towards the identification of a unigene set of barley. *Theor. Appl. Genet.* (in press).
- Millbourne, D., Meyer, R., Bradshaw, J.E., Baird, E., Bonar, N., Provan, J., Powell, W., and Waugh, R. (1997) Comparison of PCR-based marker systems for the analysis of genetic relationships in cultivated potato. *Mol. Breed.* **3**, 127-136.
- Mitchelson, K.R., Cheng, J., and Kricka, L.J. (1997) The use of capillary electrophoresis for point-mutation screening. *Trends Biotech.* **15**, 448-458.
- Miyada, C.G., and Wallace, R.B. (1987) Oligonucleotide hybridization techniques. *Methods Enzymol.* **154**, 94-107.

- Miyao, A., Zhong, H.S., Monna, L., Yano, M., Yamamoto, K., Havukkala, I., Minobe, Y., and Sasaki, T. (1996) Characterisation and genetic mapping of simple sequence repeats in the rice genome. *DNA Res.* **3**, 233-238.
- Mohan, M., Nair, S., Bhagwat, A., Krishna, T.G., Yano, M., Bhatia, C.R., and Sasaki, T. (1997) Genome mapping, molecular markers and marker-assisted selection in crop plants. *Mol. Breed.* **3**, 87-103.
- Molnar, S.J., Gupta, P.K., Fedak, G., and Wheatcroft, R. (1989) Ribosomal DNA repeat unit polymorphism in twenty-five species. *Theor. Appl. Genet.* **78**, 387-392.
- Moreau, L., Charcosset, A., Hospital, F., and Gallais, A. (1998) Marker assisted selection efficiency in populations of finite size. *Genetics* **148**, 1353-1365.
- Moreau, L., Lemarie, A., Charcosset, and Gallais, A. (2000) Economic efficiency of one cycle of marker-assisted selection. *Crop Sci.* **40**, 329-337.
- Morgante, M., and Vogel, J. (1994) Compound microsatellite primers for the detection of genetic polymorphisms. U. S. patent application no. 08/326456.
- Nakamura, Y., Leppert, M., O'Connell, P., Wolff, R., Holm, T., Culver, M., Martin, C., Fujimoto, E., Hoff, M., Kumlin, E., and White, R. (1987) Variable number of tandem repeat (VNTR) markers for human genome mapping. *Science* **235**, 516-522.
- Neuhaus, G., Werner, K., Weyen, J., Friedt, W., and Ordon, F. (2001) SNP-scanning in fragments linked to resistance genes against the barley yellow mosaic virus complex. AG Molekulare Marker in der GPZ, 25-26 Sept., 2001, Martin-Luther-Universität, Halle-Wittenberg, Germany.
- Nikiforov, T.T., Rendle, R.B., Goelet, P., Pagers, Y.H., Kotewicz, M.L., Anderson, S., Trainor, G.L., and Knapp, N.R. (1994) Genetic bit analysis: a solid phase method for typing single nucleotide polymorphisms. *Nucl. Acids Res.* **22**, 4715-4716.
- Ogihara, Y., Nemoto, Y., Murai, K., Yamazaki, Y., Shin-I., T., and Kohara, Y. (2001) Large scale analysis of ESTs in common wheat, XVIth Eucarpia congress, Plant breeding: sustaining the future: Abstract P1.23, 10th –14th Sept. 2001, Edinburgh (UK).
- Olson, M., Hood, L., Cantor, C., and Botstein, D. (1989) A common language for physical mapping of the human genome. *Science* **245**, 1434-1435.
- Orita, M., Iwahana, H., Kanazawa, H., Hayashi, K., and Sekiya, T. (1989a) Detection of polymorphisms of human DNA by gel electrophoresis as single- strand conformation polymorphisms. *Proc. Natl. Acad. Sci. USA* **86**, 2766-2770.
- Orita, M., Suzuki, Y., Sekiya, T., and Hayashi, K. (1989b) Rapid and sensitive detection of point mutations and DNA polymorphisms using the polymerase chain reaction. *Genomics* **5**, 874-879.
- Pagila, G., and Mongante, M. (1998) PCR-based multiplex DNA fingerprinting techniques for the analysis of genomes. *Mol. Breed.* **4**, 173-177.
- Pakhinyat, H., Powell, W., Baird, E., Handley, L.L., Robinson, D., Scrimgeour, C.M., Nevo, E., Hackett, C.A., Caligari, P.D.S., and Forster, B.P. (1997) AFLP variation in wild barley (*Hordeum spontaneus* C. Koch) with reference to salt tolerance and associated ecogeography. *Genome* **40**, 332-341.
- Paran, I., and Michelmore, R.W. (1993) Development of reliable PCR based markers linked to downy mildew resistance genes in lettuce. *Theor. Appl. Genet.* **85**, 985-993.
- Perrot-Minnot, M-J., Lagnel, J., Migeon, A., and Navajas, M. (2000) Tracking paternal genes with DALP markers in a pseudoarrhenotokous reproductive system : biparental transmission but haplodiploid-like inheritance in the mite *Neoseiulus californicus*. *Heredity* **84**, 702-709.
- Paul, S., Wachira, F.N., Powell, W., and Waugh, R. (1997) Diversity and genetic differentiation among populations of Indian and Kenyan tea (*Camellia sinensis* (L.) O.

- Kuntze) revealed by RFLPs, RAPDs, SSRs, and AFLPs. *Theor. Appl. Genet.* **94**, 255-263.
- Plomion, C., Hurme, P., Frigerio, J-M., Ridolfi, M., Pot, D., Pionneau, C., Avila, C., Gallardo, F., David, H., Neutelings, G., Campbell, M., Canovas, F.M., Savolainen, O., Bodenes, C., and Kerner, A. (1999) Developing SSCP markers in two *Pinus* species. *Mol. Breed.* **5**, 21-31.
- Powell, W., Morgante, M., Andre, C., Hanafey, M., Vogel, J., Tingey, S., and Rafalski, J.A. (1996) The comparison of RFLP, RAPD, AFLP and SSR (microsatellite) markers for germplasm analysis. *Mol. Breed.* **2**, 225-238.
- Prasad M., Varshney, R.K., Roy, J.K., Balyan, H.S., and Gupta, P.K. (2000) The use of microsatellites for detecting DNA polymorphism, genotype identification and genetic diversity in wheat. *Theor. Appl. Genet.* **100**, 584-592.
- Provan, J., Thomas, W.T.B., Forster, B.P., and Powell, W. (1999) *Copia*-SSR: A simple marker technique which can be used on total genomic DNA. *Genome* **42**, 363-366.
- Qi, X., Stam, P., and Lindout, P. (1998) Use of locus-specific AFLP markers to construct a high-density molecular map in barley. *Theor. Appl. Genet.* **96**, 376-384.
- Rafalski, A., Ching, A., Bhatramakki, D., Henderson, K., Jung, M., Morgante, M., Dolan, M., Register, J., Smith, O., and Tingey, S. (1999) Single nucleotide polymorphisms (SNPs) in the 3'-untranslated flanks of maize genes reveal conserved ancestral haplotypes, Cold Spring Harbor meeting on genome sequencing and biology, Cold Spring Harbor, New York.
- Rafalski, J.A., Morgante, M., Powell, W., Vogel, J.M., and Tingey, S.V. (1996) Generating and using DNA markers in plants. In: B. Birren and E. Lai (eds.), *Analysis of Non-mammalian Genomes: A Practical Guide*, Academic Press, Boca Raton, FL pp. 75-134.
- Rafalski, J.A., and Tingey, S.V. (1993) Genetic diagnostics in plant breeding: RAPDs, microsatellites and machines. *Trends Genet.* **9**, 275-280.
- Richardson, T., Cato, S., Ramser, J., Kahl, G., and Weising, K. (1995) Hybridization of microsatellites to RAPD: a new source of polymorphic markers. *Nucl. Acids Res.* **23**, 3798-3799.
- Roy, J.K., Balyan, H.S., Prasad, M., and Gupta, P.K. (2001) Use of SAMPL for a study of DNA polymorphism, genetic diversity and possible gene tagging in bread wheat. *Theor. Appl. Genet.* (in press).
- Ruan, Y., Gilmore, J., and Conner, T. (1998) Towards *Arabidopsis* genome analysis: monitoring expression profiles of 1400 genes using cDNA microarrays. *Plant J.* **15**, 821-833.
- Russell, J.R., Fuller, J.D., Macaulay, M., Hatz, B.G., Jahoor, A., Powell, W., and Waugh, R. (1997) Direct comparison of levels of genetic variation among barley accessions detected by RFLPs, AFLPs, SSRs and RAPDs. *Theor. Appl. Genet.* **95**, 714-722.
- Sapolsky, R.J. Hsie, L., Berno, A., Ghandour, G., Mittmann, M., and Fan, J.B. (1999) High throughput polymorphism screening and genotyping with high oligonucleotide arrays. *Genetic and Biomol. Engg.* **14**, 187-192.
- Scheider, K., Weisshaar, B., Borchardt, D.C., and Salamini, F. (2001) SNP frequency and allelic haplotype structure of *Beta vulgaris* expressed genes. *Mol. Breed.* **8**, 63-74.
- Schmidt, T., and Heslop-Harrison, J.S. (1996) The physical and genomic hybridization of microsatellites in sugar beet. *Proc. Natl. Acad. Sci. USA* **93**, 8761-8765.
- Schuelke, M. (2000) An economic method for the fluorescent labelling of PCR fragments. *Nature Biotechnology* **18**:233-234

- Schwengel, D.A., Jedlicka, A.E., Nanthakumar, E.J., Weber, J.L., and Levitt, R.C. (1994) Comparison of fluorescence-based semi-automated genotyping of multiple microsatellite loci with autoradiographic techniques. *Genomics* **22**, 46-54.
- Scott, K.D. (2001) Microsatellite derived from ESTs, and their comparison with those derived by other methods, in R.J. Henry (ed.), *Plant Genotyping: The DNA Fingerprinting of Plants*, CABI Publishing, Oxon, U. K. pp. 225-237.
- Scott, K.D., Eggler, P., Seaton, G., Rossetto, M., Ablett, E.M., Lee, L.S., and Henry, R.J. (2000) Analysis of SSRs derived from grape ESTs. *Theor. Appl. Genet.* **100**, 723-726.
- Scrimshaw, B.J. (1992) A simple nonradioactive procedure for visualization of (dC-dA)_n dinucleotide repeat length polymorphisms. *Bio /Techniques* **13**, 189.
- Shan, X., Blake, T.K., and Talbert, L.E. (1999) Conversion of AFLP markers to sequence-specific PCR markers in barley and wheat. *Theor. Appl. Genet.* **98**, 1072-1078.
- Somers, D.J., Kirkpatrick, R., Moniwa, M., Walsh, A., and Gustafson, J.P. (2001) Mapping wheat ESTs by haplotype specific SNPs, XVIth Eucarpia congress, Plant breeding: sustaining the future: Abstract P1.25, 10th -14th Sept. 2001, Edinburgh (UK).
- Somers, D.J., Znon, Z., Bebeli, P.J., and Gustafson, J.P. (1996) Repetitive genome-specific probes in wheat (*Triticum aestivum* L.) amplified with minisatellite core sequences. *Theor. Appl. Genet.* **93**, 982-986.
- Southern, E.M. (1975) Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J. Mol. Biol.* **98**, 503-517.
- Sreenivasulu, N., Altschmied L., Weschke, W., Panitz, R., Hähnel, U., Michalek, W., and Wobus, U. (2001a) Identification of genes specifically expressed in maternal and filial tissues of barley caryopsis: a cDNA array analysis. *Mol. Gen. Gen.* (in press).
- Sreenivasulu, N., Wobus, U., Miranda, M., and Weschke, W. (2001b) A set of salt-induced genes identified in tolerant foxtail millet seedlings by using a barley cDNA macroarray, 4th International triticeae symposium, 10th-12th Sept. 2001, Cordoba, Spain, pp. 137.
- Synvanen, A.-C., Aalto-Setälä, K., Harju, L., Kontula, K., and Soderlund, H. (1990) A primer-guided nucleotide incorporation assay in the genotyping of apolipoprotein E. *Genomics* **8**, 684-692.
- Tautz, D. (1989) Hypervariability of simple sequences as a general source for polymorphic DNA markers. *Nucl. Acids Res.* **17**, 6443-6471.
- Tyagi, S., Bratu, D.P., and Kramer, F.R. (1998) Multicolour molecular beacons for allele discrimination. *Nature Biotech.* **16**, 49-53.
- Tyagi, S., and Kramer, F.R. (1996) Molecular beacons: probes that fluoresce upon hybridization. *Nature Biotech.* **14**, 303-308.
- vanEck, H.J., van der Voort, R., Draaistra, J., van Zandvoort, P., van Enckevort, E., Seggers, B., Peleman, Jacobson, E., Helder, J., and Bakker, J. (1995) The inheritance and chromosomal localization of AFLP markers in a non-inbred potato offspring. *Mol. Breed.* **1**, 397-410.
- Varghese, J.P., Rudolph, B., Uzunova, M.I., and Ecker, W. (2000) Use of 5'-anchored primers for the enhanced recovery of specific microsatellite markers in *Brassica napus* L. *Theor. Appl. Genet.* **101**, 115-119.
- Varshney, R.K., Sharma, P.C., Gupta, P.K., Balyan, H.S., Ramesh, B., Roy, J.K., Kumar, A., and Sen, A. (1998) Low level of polymorphism detected by SSR probes in bread wheat. *Plant Breed.* **117**, 182-184.
- Vivek, B.S., and Simon, P.W. (1999) Linkage relationships among molecular markers and storage root traits of carrot (*Daucus carota* L. ssp. *sativus*). *Theor. Appl. Genet.* **99**, 58-64.

- Vos, P., Hogers, R., Bleeker, R., Reijans, M., van de Lee, T., Hornes, M., Frijters, A., Pot, J., Peleman, J., Kupier, M., and Zabeau, M. (1995) AFLP: a new technique for DNA fingerprinting. *Nucl. Acids Res.* **23**, 4407-4414.
- Vosman, B., Arens, P., Rus-Kortekaas, W., and Smulders, M.I.M. (1992) Identification of highly polymorphic DNA regions in tomato. *Theor. Appl. Genet.* **85**, 239-244.
- Wang, D.G., Fan, J.B., Siao, C.J., Berne, A., Young, P., Sapolsky, R., Ghandour, G., Perkins, N., Winchester, E., Spencer, J., Kruglyak, L., Stein, L., Hsie, L., Topaloglou, T., Hubbell, E., Robison, E., Mittmann, M., Morris, M.S., Shen, N., Kilbum, D., Riox, J., Nusbaum, C., Rozen, S., Hudson, T.J., Lipshutz, R., Chee, M., and Lander, E.S. (1998) A large-scale identification, mapping, and genotyping of single nucleotide polymorphisms in the human genome. *Science* **280**, 1077-1082.
- Waugh, R., McLean, K., Flavell, A.J., Pearce, S.R., Kumar, A., Thomas, B.B.T., and Powell, W. (1997) Genetic distribution of *BARE-1*-like retrotransposable elements in the barley genome revealed by sequence-specific amplification polymorphisms (S-SAP). *Mol. Gen. Genet.* **255**, 687-694.
- Welsh, J., and McClelland, M. (1990) Fingerprinting genomes using PCR with arbitrary primers. *Nucl. Acids Res.* **18**, 7213-7218.
- Whitkus, R., Doebley, J., and Jonathan, F. (1994) Nuclear DNA markers in systematics and evolution, in R.L. Phillips and I.K. Vasil (eds.), *DNA- Based Markers in Plants*, Kluwer Academic Publishers, Dordrecht, pp. 116-142.
- Whittaker, J.C., Curnow, R.N., Haley, C.S., and Thompson, R. (1995) Using marker-maps in marker-assisted selection. *Genetical Res.* **66**, 255-265.
- Whittaker, J.C., Haley, C.S., and Thompson, R. (1997) Optimal weighting on information in marker-assisted selection. *Genetical Res.* **69**, 137-144.
- Williams, J.G.K., Kubelik, A.R.K., Livak, J.L., Rafalski, J.A., and Tingey, S.V. (1990) DNA polymorphisms amplified by random primers are useful as genetic markers. *Nucl. Acids Res.* **18**, 6531-6535.
- Witsenboer, H., Vogel, J., and Michelmore, R.W. (1997) Identification, genetic localization, and allelic diversity of selectively amplified microsatellite polymorphic loci in lettuce and wild relatives (*Lactuca* spp.). *Genome* **40**, 923-936.
- Wolff, K., Zietkiewicz, E., and Hofstra, H. (1995) Identification of chrysanthemum cultivars and stability of fingerprint patterns. *Theor. Appl. Genet.* **91**, 439-447.
- Wolf, M., Kota, R., Thiel, T., Michalek, W., and Graner, A. (2001) ESTs as a resource for the systematic development of novel markers in barley, AG Molekulare Marker in der GPZ, 25-26 September, 2001, Martin-Luther-Universität, Halle-Wittenberg, Germany.
- Wu, D.Y., Ugozzoli, L., Pal, B.K. and Wallace, R.B. (1989) Allele specific enzymatic amplification of β -globin genomic DNA for diagnosis of sickle cell anaemia. *Proc. Natl. Acad. Sci. U.S.A.* **86**, 2757-2760.
- Wu, K., Jones, R., Dannaeburger, L., and Scolnik, P.A. (1994) Detection of microsatellite polymorphisms without cloning. *Nucl. Acids Res.* **22**, 3257-3258.
- Xie, C., and Xu, S. (1998) Efficiency of multi stage marker-assisted selection in the improvement of multiple quantitative traits. *Heredity* **80**, 489-498.
- Xu, R-Q., Tomooka, N., and Vaughan, A. (2000) AFLP markers for characterizing the Azuki bean complex. *Crop Sci.* **40**, 808-815.
- Yamamoto, K., and Sasaki, T. (1997) Large- scale EST sequencing in rice. *Plant Mol. Biol.* **35**, 135-144.

- Young, N.D., Denny, R.L., Concibido, V.C., Lange, D.A., Orf, J.H. (1995) Marker assisted breeding in practice: RFLPs and soybean cyst nematode resistance, *Induced Mutations and Molecular Techniques for Crop Improvement*, IAEA, Vienna, pp. 245-251.
- Young, N.D. (1999) A cautiously optimistic vision for marker-assisted breeding. *Mol. Breed.* **5**, 505-510.
- Zhang, W., and Smith, C. (1992) Computer simulation of marker-assisted selection utilizing linkage disequilibrium. *Theor. Appl. Genet.* **83**, 813-820.
- Zhang, W., and Smith, C. (1993) Simulation of marker-assisted selection utilizing linkage disequilibrium: the effect of several additional factors. *Theor. Appl. Genet.* **86**, 492-496.
- Ziegle, J.S., Su, Y., Corcoran, K.P., Nie, L., Mayrand, P.E., Hoff, L.B., McBride, L.B., Kronoick, M.N., and Diehl, S.R. (1992) Application of automated DNA sizing technology for genotyping microsatellite loci. *Genomics* **14**, 1026-1031.
- Zietkiewicz, E., Rafalski, A., and Labuda, D. (1994) Genome fingerprinting by simple sequence repeat (SSR)-anchored polymerase chain reaction amplification. *Genomics* **20**, 176-183.
- Zhu, J., Gale, M.D., Quarrie, S., Jackson, M.T., and Bryan, G.J. (1998) AFLP markers for the study of rice diversity. *Theor. Appl. Genet.* **96**, 602-611.

3

MOLECULAR MARKER ASSISTED BREEDING

DARSHAN S. BRAR

International Rice Research Institute

Los Baños, Laguna, Philippines

INTRODUCTION

Molecular markers have ushered in a new era in genetics and crop improvement. These markers are becoming powerful tools to develop dense molecular genetic maps, understanding genetic architecture of crop plant genomes, mapping of agronomically important genes including quantitative trait loci (QTLs), marker assisted selection (MAS), map based cloning of useful genes, determining genetic diversity in crop germplasm and pathogen populations, precise monitoring of alien introgression and in analysis of evolutionary relationships.

Numerous genes of agronomic importance particularly for resistance to diseases and insects have been tagged with molecular markers in several crop plants. Tight linkage between a marker and a trait has led to the successful transfer of genes through MAS. Notable example includes pyramiding of genes *Xa4*, *xa5*, *xa13*, *Xa21* for bacterial blight resistance in rice. It has become possible to monitor mapped QTL regions through MAS in developing near isogenic lines for QTLs. There is an urgent need to identify molecular markers tightly linked with the genes governing resistance to major diseases and insects, tolerance to abiotic stresses, quality and other agronomic traits. PCR-based markers and SNPs should be the markers of choice in MAS. Emphasis should be placed on molecular mapping and introgression of favorable QTLs into elite breeding lines of crops particularly those governing tolerance to abiotic stresses and yield components having larger effects over environments. Mapping and introgression of complementary yield enhancing loci/QTL from the exotic or

wild species germplasm should be explored to further enhance yield potential of crops. Fine mapping and cloning of QTL adds another challenge to molecular marker technology aimed at crop improvement. Intergenomic cloning of orthologous genes should be explored. DNA based markers are expected to serve as good descriptors in varietal identification and in efficient management and utilization of genetic resources. Molecular markers offer unique advantage to develop overlapping substitution lines with defined chromosome segments of donor parents which have become important genetic resource for gene discovery and functional genomics. Continued efforts should be made to make molecular marker assisted breeding cost effective and integral part of conventional plant breeding programs.

The discovery of Mendel's laws of inheritance in 1901 and chromosomal theory of inheritance in 1902 laid the foundation of classical genetics. These discoveries led to the concept that characters are controlled by factors (now called genes), which are inherited from parents to the progenies. Further, chromosomes are the main vehicles of heredity. The first genetic map was developed in 1913 by A.H. Sturtevant, which consisted of 5 genes located in a linear fashion on a chromosome of *Drosophila*. Since then genetic maps have been developed in several plant species. These maps whether based on genetic recombination, restriction analysis, physical location or DNA sequence are predicted on Sturtevant's logical notion that gene order in a chromosome could be displayed as a linear array of genetic markers.

Three kinds of markers, morphological (plant traits), biochemical (proteins and isozymes), and molecular (DNA) have been used in construction of genetic maps. Morphological markers are limited in number, are influenced by the environments, are development stage specific and some have pleiotropic effect; hence are not suitable for genetic mapping. Isozymes markers are also limited in number and thus cannot be used to prepare saturated genetic maps.

Molecular markers commonly referred as DNA markers are numerous in number and their discovery represents a milestone in genetics as they provide the capacity for complete coverage of crop genome. These markers show Mendelian inheritance, are stably inherited, have no pleiotropic effect and are

unaffected by the environment and express at all developmental stages. A number of reviews have been published on molecular markers and their application in crop improvement (Tanksley 1983, 1993, Paterson et al. 1994, Lee 1995, Mohan et al. 1997, Brar and Dhaliwal 1997, Khush and Brar 1998, Mackill and Ni 2001, McCouch et al. 2001). Majority of DNA polymorphisms are selectively neutral. The genetic variation can be the result of a simple point mutation, DNA insertion/deletion event or change in repeat copy number at some hypervariable DNA or microsatellite. The methods of detection of polymorphism involve the use of restriction endonuclease, nucleic acid hybridization, or DNA sequence amplification.

Molecular markers are of great value in applying genetic technologies to crop improvement such as determining genetic diversity, marker assisted selection, gene-pyramiding, QTL mapping, map-based cloning of important genes, monitoring introgression from exotic and wild species germplasm, DNA fingerprinting of crop germplasm and pathogen populations.

MARKER ASSISTED SELECTION

The potential value of genetic markers, linkage groups and their association with agronomic traits has been known for more than 80 years. The usefulness of marker assisted selection (MAS) was recognized as early as 1923 when Sax demonstrated the association of seed size with seed coat pigmentation in beans.

The first RFLP map in crop plants was constructed in 1986 with only 57 loci (Bernatzky and Tanksley 1986). Subsequent research led to the construction of a series of molecular maps in several crop plants. Such high density molecular genetic maps made it possible to apply MAS originally proposed by Sax (1923) and Thoday in 1961, in plant breeding. The concept for selecting desirable lines based on genotype rather than phenotype created strong interest among plant breeders worldwide.

Following are some of the major components of MAS aimed at enhancing the efficiency of plant breeding.

1. Types of molecular markers used

2. High density molecular genetic maps
3. Gene tagging through tight linkage with molecular markers
4. MAS protocols

TYPES OF MOLECULAR MARKERS

Several kinds of molecular markers are available namely: Restriction fragment length polymorphism (RFLP), Random amplified polymorphic DNA (RAPD), Amplified fragment length polymorphism (AFLP), Amplicon length polymorphism (ALP), DNA amplification fingerprinting (DAF), cleaved amplified polymorphism (CAP), sequence characterized amplified region (SCAR), simple sequence repeats (SSR), sequenced tagged sites (STS), expressed sequence tag (EST), and single nucleotide polymorphism (SNP), etc.

Among these, PCR based markers such as SSR are the markers of choice. When genome sequence of major crops become known, the SNP markers are likely to become important in genetic and breeding research aimed at crop improvement.

RFLP: RFLPs have been important markers in genetic mapping. These have been used to generate saturated genetic maps in several crops such as tomato, rice, wheat, maize, lettuce, barley, etc. RFLP analysis can make use of single-copy probes or repetitive probes. Both DNA and random genomic libraries have been used as sources of probes for RFLP mapping. Most laboratories have used probes radioactively labeled with P^{32} . Non-radioactively labeled probes can also be used in RFLP analysis, and are desirable for safety purposes and in situations where isotopes are difficult to obtain or use.

RAPD: Conventional RFLP analysis is limited by the relatively large amount of DNA required, Southern blotting and hybridization, need for radioactive isotopes and autoradiography. These factors make conventional RFLP analysis relatively slow and expensive. The development of new methods based on PCR amplification of genomic DNA and detection of differences in the PCR products have become popular. Genetic tests based on PCR are simple to perform. These DNA markers are generated through the use of the polymerase chain reaction (PCR). Using PCR, particular fragments of DNA are amplified in large quantities, considerably facilitating their detection. The amplified portions of

DNA are defined by short, (16-20 base) synthetic DNA molecules (primers) that match the ends of the DNA to be amplified. The PCR-based markers can be utilized in much the same way that RFLPs are used. In some cases RFLP markers are actually converted to PCR-based markers.

AFLP: Amplified fragment length polymorphism (AFLP) are DNA fragments with different nucleotide sequences, of which large number of copies have been synthesized (amplified) via PCR. The procedure combines the restriction site variation similar to that sampled in RFLPs with the exponential amplification of PCR based marker systems. Only the DNA sequence interval (several thousand nucleotides) between the sites where one or more base pairs or polynucleotide (15 to 35 bp) or oligonucleotide (2 to 10 bp) primers anneal to a DNA template is amplified. Like RFLPs, the AFLPs are highly heritable, polymorphic, apparently selectively neutral, and are nearly ubiquitous in plant tissue. The technique is based on selective PCR amplification of restriction fragments from a total digest of genomic DNA. It involves (i) restriction of DNA and ligation of the oligonucleotide adapters (ii) PCR amplification of sets of restriction fragments and (iii) gel analysis of the amplified fragments. AFLPs are revealed through fractionation by agarose or polyacrylamide gels (PAGE), the amplified DNA is generally stained with ethidium bromide, silver salts or other non-specific stain for DNA. Typically 60-100 restriction fragments are amplified and detected on denaturing polyacrylamide gels. The technique has the distinct advantage of detection of a large number of polymorphisms from a single PCR reaction without prior sequence information.

Microsatellite: These are a class of repeat sequences that are comprised of tandem repeats of short, core sequences dispersed throughout the genome. When these core sequences are hybridized to restriction enzyme digested genomic DNA, they detect several hyper variable loci simultaneously. Microsatellites are quite attractive as the assay is PCR based. It is sufficient to merely separate the amplification products by electrophoresis. If the DNA sequence information is made available, individual laboratories can synthesize their own oligonucleotide.

Sequence tagged sites (STS): STS is a short stretch of genomic sequence that can be detected by PCR. Each STS is mapped to a specific site as landmark in

the genome. In human and animal genome research, a large number of STS have been produced by analyzing RFLP markers, micro satellite, expressed sequences, and fragments of yeast artificial chromosome (YAC) inserts. STS is important to convert a genetic map to a physical map. Based on sequence information of STS, specific genes can also be isolated.

Expressed sequence tag (EST): EST markers have been developed in several laboratories. These markers are developed based on extensive sequence data of regions of the genome which are expressed. However, once developed they provide high quality and highly consistent results. These markers are directly associated with functional genes. EST markers are likely to be less polymorphic than SSR markers.

Single nucleotide polymorphisms (SNP): The vast majority of differences between individuals are point mutations due to SNP. As such, there are a vast number of potential SNP markers in all species. Considerable amounts of sequence data are required from parental genotypes to develop SNP markers, however, their great advantage lies in the potential to screen them using methods which do not involve electrophoresis, such as microarrays.

MOLECULAR GENETIC MAPS

Since the first RFLP map constructed by Bernatzky and Tanksley in 1986 in tomato with 57 loci, series of molecular genetic maps have been developed in rice, wheat, tomato, maize, barley, lettuce, *Brassica* and more recently in many horticultural and tree species. Consistent efforts have resulted in the development of high density molecular maps in many crops. As an example, molecular map of rice consists of 2300 DNA markers (Harushima et al. 1998). Such high density maps have been used in tagging the genes governing important agronomic traits with molecular markers. Subsequently, protocols for MAS have been developed (Zheng et al. 1995) and applied in plant breeding programs.

GENE TAGGING

The process of locating genes of interest via linkage to molecular markers is referred to as gene tagging. Classical "phenotype" and molecular marker

"genotypes" must be evaluated on the same individuals (or related progenies) and the data analyzed to determine if any of the markers co-segregate with the target phenotype. A molecular marker very closely linked to a gene can act as a "tag" which can be used for indirect selection of the gene in breeding programs. It is often desirable to 'tag' an agronomically valuable gene with closely linked molecular markers. One way to accomplish this is to map the trait relative to a set of molecular markers in segregating populations.

A large number of genes for disease and insect resistance and for several other characters have been tagged with molecular markers in rice (Table 1). Several agronomically important genes have been tagged with molecular markers in many other crops such as wheat, maize, tomato, soybean, cotton, etc. (Mohan et al. 1997, Yenco et al. 2000). Notable examples where tight linkages have been detected between genes and molecular markers and are suitable to apply MAS include SSR marker with *Rcs3* gene governing frogeye leaf spot resistance in soybean (Rouf Mian et al. 1999); soybean cyst nematode resistance for the loci located in linkage 'G' of soybean (Concibido et al. 1996); *Tm-2* for resistance to tobacco mosaic virus with TG101 marker (Young and Tanksley 1989); *Cf9* gene for resistance to *Cladosporium fulvum* with TG236 (Thomas et al. 1995). Eighteen RAPD markers were linked to 11 genes (*H3*, *H5*, *H6*, *H9*, *H10*, *H11*, *H12*, *H13*, *H14*, *H16* or *H17*) for Hessian fly resistance (Dweikat et al. 1997); and complete linkage of a RAPD marker with *bc-1²* gene for common mosaic virus resistance in bean (Miklas et al. 2000). Markers (umc28, csu68 and umc62) linked with apomictic mode of reproduction have been identified in maize (Leblanc et al. 1995). Similarly Ozias-Akins et al. (1993) identified linkage of three molecular markers with aposporous mode of reproduction in *Pennisetum*. These probes UGT197, UGT184 and UGT-1 were converted to STS. A 12.4 kb fragment amplified by RAPD primer OPK18 was found associated with resistance of HD29 variety of *Triticum aestivum* to Karnal bunt in bulk segregant analysis from HD29xWL711 cross (Singh and Dhaliwal, Per. Comm.). Mohan et al. (1997), Yenco et al. (2000) and Mackill and Ni (2001) have reviewed molecular tagging of different genes governing agronomically

Table 1. Some examples on mapping of genes of agronomic importance with molecular markers in rice.

Gene	Trait	Chromosome	Linked marker and distance (cM)	
<i>Pi1</i>	Blast resistance	11	Npb181	3.6
<i>Pi2</i>	Blast resistance	6	RG64	2.8
<i>Pi4</i>	Blast resistance	12	RG869	15.3
	Blast resistance		RZ536	7.9
<i>Piz5</i>	Blast resistance	6	RG64	2.1
<i>Pita</i>	Blast resistance	12	RZ397	3.3
	Blast resistance		RG241	5.2
<i>Pi5(t)</i>	Blast resistance	4	RG498	5-10
	Blast resistance		RG7BB	
<i>Pi6</i>	Blast resistance	12	RG869	20
<i>Pi7</i>	Blast resistance	11	RG103	5-10
<i>Pi9</i>	Blast resistance	6	RG16	-
<i>Pi10</i>	Blast resistance	5	RRF6, RRH18	-
<i>Pi11</i>	Blast resistance	8	BP127	2.4
<i>Pi12</i>	Blast resistance	12	RG869	5.1
<i>Pi lm2</i>	Blast resistance	11	RZ536	6.0
<i>Pib</i>	Blast resistance	2	RZ'23	
<i>Xa1</i>	Bacterial blight resistance	4	Npb235	3.3
<i>Xa2</i>	Bacterial blight resistance	4	Npb235	3.4
<i>Xa3</i>	Bacterial blight resistance	11	Npb181	2.3
			Npb78	3.5
<i>Xa4</i>	Bacterial blight resistance	11	Npb181	1.7
<i>xa5</i>	Bacterial blight resistance	5	RG556	0.1
<i>Xa10</i>	Bacterial blight resistance	8	OP07-2000	5.3
			RG136	3.8
<i>xa13</i>	Bacterial blight resistance	8	RG136	3.8
<i>Xa21</i>	Bacterial blight resistance	11	PTA248	0.1
			RG103	
<i>Stvb1</i>	Rice stripe virus resistance	11	XNpb220	0.9
RTSV	Rice tungro spherical virus resistance	4	RZ262	4.5
HbV	hojablanca resistance	12	-	-
<i>Bph1</i>	Brown planthopper resistance	12	Npb248	-
<i>bph2</i>	Brown planthopper resistance	12	-	3.5
<i>Bph10</i>	Brown planthopper resistance	12	RG457	3.7
<i>Qbp1</i>	Brown planthopper resistance	3	RG2443	14.6
<i>Qbp2</i>	Brown planthopper resistance	4	Y3653R	0.4

Table 1. Continued.

Gene	Trait	Chromosome	Linked marker and distance (cM)	
<i>Gm1</i>	Gall midge resistance	-	OPN-15	-
<i>Gm2</i>	Gall midge resistance	4	RG329	1.3
<i>gm3</i>	Gall midge resistance	-	OP4-01	-
<i>Gm4</i>	Gall midge resistance	8	E20-570	-
<i>Gm5</i>	Gall midge resistance	-	OPE-01	-
<i>Gm6(t)</i>	Gall midge resistance	4	RG214	1.0
<i>Glh</i>	Green leaf hopper resistance	4	RZ262	2.1
<i>Rf</i>	Fertility restoration	10	RG561, CD094	
<i>Rf2</i>	Fertility restoration	1	RZ382-RG458	
<i>Rf3</i>	Fertility restoration	1	RG532	0-2
<i>S5</i>	Wide compatibility	6	RG213	4.4
<i>tms1</i>	Thermosensitive male sterility	8	-	-
<i>tms3</i>	Thermosensitive male sterility	8	OPAC3-640	7.7
<i>pms1</i>	Photoperiod sensitivity	7	RG477	4.3
	male sterility			
<i>pms2</i>	Photoperiod sensitive	3	RG191	-
	male sterility			
<i>Se1</i>	Photoperiod sensitivity	6	RG64	0
<i>Se3</i>	Photoperiod sensitivity	6	A18	5-10
<i>sdg(t)</i>	Semidwarf	5	RZ182	4.3
<i>sd1</i>	Semidwarf	1	RG109	0.8
<i>fgr</i>	Fragrance	8	RG28	4.5
<i>Sub-t</i>	Submergence tolerance	3	RZ698	-

Modified from Khush and Brar (1998)

important traits in crop plants. Traits closely linked with molecular markers are increasingly transferred into elite breeding lines of crop plants through MAS.

More efficient ways of obtaining linked markers utilize near isogenic lines (NILs). NILs have been produced for a variety of traits in several crop plants by repeated cycles of backcrossing and selection. Michelmore et al. (1991) devised 'bulked segregant analysis' for tagging of genes with molecular markers. The F_2 plants from a lettuce population segregating for resistance to downy mildew were separated into two groups those which were homozygous resistant and those which were homozygous for sensitivity (heterozygotes previously identified by progeny tests were excluded). DNA was then combined from several plants from each group to form two DNA pools. RAPD primers were then used to amplify DNA from each of the two pools. Three RAPD polymorphisms were identified and all were linked to the resistance gene. The bulk segregant analysis is quite popular in various gene tagging experiments.

MARKER AIDED SELECTION (MAS) PROTOCOLS AND THEIR APPLICATION

The development of saturated molecular maps and tight linkage of target genes with the molecular markers and conversion of these markers to PCR based markers have made it feasible to use MAS in plant breeding programs. In MAS individuals carrying target genes are selected in a segregating population based on patterns of tightly linked markers rather than on their phenotypes. Thus, the population can be screened at any growth stage and in various environments. MAS can overcome interference from interactions between alleles of a locus or other loci. MAS can increase the efficiency and accuracy of selection, especially for traits that are difficult to phenotype. Frisch et al. (1999a, b) have reviewed the required population size in MAS. Young (1999) reviewed the usefulness of MAS keeping in view the cost involved, trait under consideration and availability of mapping populations and marker systems. There is a need to improve the experimental designs and integration of molecular marker data with the genomics and bioinformatics for enhancing the efficiency of MAS.

REQUIREMENTS FOR MAS

1. Marker should co-segregate or be closely linked (1 cM or less) with the desired trait.
2. Feasibility to screen large populations with the marker preferable PCR based markers.
3. The screening technique should have high reproducibility.
4. The technique should be cost effective.

PYRAMIDING GENES THROUGH MAS

Tight linkage of molecular markers with the target gene(s) has led to the application of MAS in breeding programs. During the last decade, numerous genes in several crops have been tagged with molecular markers which could be used advantageously in MAS. MAS protocols have been well established in rice, tomato, wheat, and barley. Some notable examples on the successful use of MAS in pyramiding of genes include bacterial blight (BB) resistance (Yoshimura et al. 1995, Huang et al. 1997, Sanchez et al. 2000, Singh et al. 2001), blast resistance (Hittalmani et al. 2000) and gall midge resistance in rice (Katiyar et al. 2001). In the breeding programme to combine genes for BB resistance, the parents in each cross carry either *xa13* or *Xa21*. *Xa21* confers resistance to all six Philippine races of BB whereas *xa13* confers resistance to race 6 only. Since the effect of *xa13* is masked by *Xa21*, it is difficult to select plants carrying both genes by conventional method alone. Using the marker RG136 which is closely linked to *xa13*, identification of plants carrying both genes was facilitated. This illustrates the usefulness of DNA markers in selecting for recessive genes such as *xa13* where the presence of the gene in the heterozygous condition cannot be detected without progeny testing or in cases where different genes have the same effect on a character and their genotypes cannot be readily identified through conventional approaches. Huang et al. (1997) successfully pyramided four genes for BB resistance (*Xa4*, *xa5*, *xa13*, *Xa21*). Breeding lines with two, three and four genes were developed. The pyramided lines showed wider spectrum of resistance than lines with only a single gene. Yoshimura et al. (1995) selected lines carrying *Xa4* + *xa5* and *Xa4* + *Xa10* using RFLP and RAPD markers linked to the BB resistance genes. Lines carrying *Xa4* + *xa5* had higher resistance to isolate of race 4 than were

either of the parental lines. Such pyramided lines are useful for developing varieties with durable resistance.

The STS markers were used to pyramid three genes (*xa5*, *xa13*, *Xa21*) into elite breeding lines of new plant type rice (Sanchez et al. 2000). The STS marker used and the primary sequence are given in Table 2. STS for the two resistance genes were developed based on DNA sequences of linked RG556 and RG207 for *xa5* and RG136 for *xa13*. The STS markers for *Xa21* was from the sequence of a genomic clone RAPD248. BC₃F₃ NIL were developed through MAS. The BC₃F₂ plants having 2 or 3 genes showed a wider spectrum and increased level of resistance to the BB than those having a single BB resistance gene (Table 3). Progeny tests of F₃ lines showed that MAS had an accuracy of 95 and 96% for identifying homozygous resistant plants for *xa5* and *xa13* respectively (Table 4). These results on pyramided lines in three genetic backgrounds (IR24, PR106 and new plant type) demonstrate the usefulness of MAS in pyramiding genes for BB resistance, particularly *xa5* and *xa13* that are difficult to select through conventional breeding in the presence of a dominant *Xa21* (Huang et al. 1997, Sanchez et al. 2000, Singh et al. 2001). Singh et al. (2001) pyramided three genes for BB resistance (*xa5*, *xa13* and *Xa21*) into an indica cultivar PR106 through MAS. Pyramided lines were evaluated with 17 isolates of *Xanthomonas oryzae* from Punjab, India and six races of BB from the Philippines. The genes in combination provided a wide spectrum of resistance to the pathogen than the same gene present singly.

In order to improve the durability of resistance, molecular markers have been used to identify pathogen populations with wide diversity. Such populations are used to screen genotypes resistant to BB. Various deployment strategies are used based on understanding of pathogen population genetics and on the genetic basis of durable resistance (Leung et al. 1993).

Hittalmani et al. (2000) fine mapped three genes (*Pi1*, *Piz-5*, *Pita*) for blast resistance in rice. These genes were pyramided through MAS. The lines carrying two and three gene combinations when tested in Philippines and India showed that genes in combinations have enhanced resistance than when present singly. PCR based markers have shown linkage with the gall midge resistance

Table 2. Sequence tagged site markers used in marker-assisted selection of resistance genes to *Xanthomonas oryzae* pv. *oryzae* in rice

Marker	Chromosome	Primer sequence	Enzyme	Linked gene	Distance (cM)
RG556	5	5' TAG CTG CTG CCG TGC TGT GC 3' 5' AAT ATT TCA GTG TGC ATC TC 3'	<i>Mae</i> II	<i>xa5</i>	0.0
RG207	5	5' ATT GTT ACG TTT GGT GGG GG 3' 5' GCC ATG GCG ACT GTC AGT CG 3'	-	<i>xa5</i>	0.0
RG136	8	5' TCC CAG AAA GCT ACT ACA GC 3' 5' GCA GAC TCC AGT TTG ACT TC 3'	<i>Hinf</i> I	<i>xa13</i>	4.9
pTA 248	11	5' AGACGCGGAAGGTGGTTCCCGGA 3' 5' AGACCGGTAAATCGAAAGATGAAA 3'	-	<i>Xa21</i>	0.0

Sanchez et al. 2000 (with permission from Crop Science Society of America).

Table 3. BB reaction of selected BC₃F₂ plants derived from the cross of IRBB59 and two elite breeding lines of new plant type rice, IR65600-42 and IR65598-112 selected through MAS.

BC ₃ F ₂ plants	Recurrent parent	Target genes			Reactions to Xoo races					
		xa5	xa13	Xa21	1 (PXO61)	2 (PXO85)	3 (PXO79)	4 (PXO71)	5 (PXO112)	6 (PXO99)
-	IR65600-42				MS	MS	MS	MS	S	S
-	IR65598-112				S	S	S	S	S	S
-	IRBB59 (donor)	+	+	+	R	R	R	R	R	R
1	IR65598-112	+			R	R	R	R	R	S
2	IR65598-112		+		S	S	S	S	S	R
3	IR65598-112			+	R	R	R	R	R	R
4	IR65598-112	+	+		R	R	R	R	R	R
5	IR65598-112	+		+	R	R	R	R	R	R
6	IR65598-112	+		+	R	R	R	R	R	R
7	IR65598-112		+	+	R	R	R	R	R	R
8	IR65598-112		+	+	R	R	R	R	R	R
9	IR65598-112	+	+	+	R	R	R	R	R	R
10	IR65598-112	+	+	+	R	R	R	R	R	R
11	IR65600-42	+			R	R	R	R	R	S
12	IR65600-42		+	+	R	R	R	R	R	R
13	IR65600-42	+		+	R	R	R	R	R	R
14	IR65600-42		+	+	R	R	R	R	R	R
15	IR65600-42		+	+	S	S	S	S	S	R

Modified from Sanchez et al. 2000

*Presence of the gene was inferred through linked sequence tagged site markers; R = resistant, S = susceptible; MS = moderately susceptible

Table 4. Accuracy of marker-assisted selection for the resistance genes, *xa5* and *xa13*, based on STS marker verified by F₃ progeny testing.

F ₃ plants		No. of BC ₃ F ₃ families with each resistant phenotype*			Accuracy (%)
Marker genotype	No.	RR	Rr	rr	
(<i>xa5</i>) RR	23	22	1	0	95.7
(<i>xa5</i>) Rr	69	8	60	1	89.4
(<i>xa5</i>) rr	20	0	1	19	95.0
(<i>xa13</i>) RR	19	17	2	0	80.5
(<i>xa13</i>) Rr	40	2	35	3	87.5
(<i>xa13</i>) rr	24	0	1	23	95.8

*RR = homozygous susceptible; Rr = heterozygotes; rr = homozygous resistant.

Sanchez et al. 2000. (with permission Crop Science Society of America)

genes (*Gm1*, *Gm2*, *gm3*, *Gm4*, *Gm5*, *Gm6(t)*) (Table 1). Katiyar et al. (2001) developed a PCR based marker assisted selection kit for transferring *Gm6(t)* gene into susceptible cultivars of China. The kit contains PCR primer pairs on the terminal sequences of the RG214 and RG276 probes. Penner et al. (1995) converted a RFLP marker associated with stem rust resistance in barley to an allele specific PCR marker useful for MAS.

Pyramiding epistatic resistance genes through conventional breeding has been difficult. As discussed in the previous section, MAS using tightly linked markers could be used to pyramid epistatic genes. Miklas et al. (2000) used near isogenic lines and identified RAPD markers completely linked with *bc-1²*. The gene *bc-1²* is masked by *bc-2²* and *bc-3³* genes. To facilitate MAS, the RAPD marker has been converted to a SCAR marker, SBD5₁₃₀₀. Tight linkage was confirmed between SBD5₁₃₀₀ and *bc-1²* gene.

MAS FOR QTL

MAS also provides new opportunities to transfer and combine QTLs into agronomically desirable genotypes. A number of agronomically important characters such as yield, yield components, tolerance to abiotic stresses and quality are of quantitative nature. These are controlled by a relatively large number of loci, each of which can make a small positive or negative contribution to the final phenotypic value of the traits. Such loci are termed 'quantitative trait loci' 'QTLs'. The genes governing such traits called polygenes or minor genes also follow Mendelian inheritance but are greatly influenced by the environments. Biometrical procedures involving special experimental designs and data analyses are used to study genetics of quantitative traits.

Molecular markers are of great value to introgress the QTLs identified in one parent to an array of elite breeding lines. This would be an important component of marker assisted breeding which otherwise is difficult to achieve through conventional procedures. However, precise location of QTLs applicable over range of environments and genotypes is essential before these QTLs can be used in MAS.

As early as 1923, Sax reported the association of seed size with seed-coat pigmentation. Later, Thoday (1961) proposed the use of linkage of single gene markers with polygenes. However, such markers were few, and were affected by the environment. The advent of molecular markers have made it possible to map the QTL having large genotypic effect on phenotype.

Paterson et al. (1988) used RFLP linkage map of tomato to resolve QTL into discrete Mendelian factors and located 6 QTL controlling fruit mass, 4 QTL for soluble solid concentration and 5 QTL for fruit pH in tomato. Since then, QTLs have been mapped for several traits in many crop plants. Methodologies for detecting and mapping of QTL involving different population sampling strategies, type of populations, threshold levels for detecting QTL and missing data have been described in several publications (Paterson et al., 1988, 1998, Li 2001). The most important fact is uncertainty of QTL location relative to DNA markers. Some limitations include lack of recombinant gametes and genetic heterogeneity for regional restrictions to recombination, missing data for markers and traits, linked QTL and probably QTL of truly minor effect. Xu (1997) has reviewed different methodologies for mapping and pyramiding of QTLs through MAS.

Ribaut and Betran (1999) proposed a new approach for single large scale marker-assisted selection (SLS-MAS). In this approach, plants are selected in early generation with a fixed, favorable genetic background at specific loci, conducting a large scale MAS while maintaining as much as possible the allelic segregation in the rest of the genome. After selection with DNA markers, the genetic diversity at unselected loci allows breeders to generate new varieties and hybrids through conventional breeding. The scheme provides flexibility and has been used in maize for polygenic traits. However, additional data are awaited for large- scale application in other programs.

Wang et al. (1994) in rice located two major genes, *Pi5(t)*, *Pi7(t)* on chromosomes 4 and 11 respectively. Nine QTLs having quantitative effects on resistance to isolate PO6-6 (blast resistance) were also identified.

Stubber et al. (1992) mapped QTLs and evaluated their phenotypic effects associated with seven agronomic traits in a cross involving two elite maize inbred lines. Phenotypic evaluations were made on nearly 100,000 plants grown in three states (North Carolina, Iowa and Illinois). Although the experiment was conducted in six different environments, however there was little evidence for G x E interaction for most QTLs. Paterson et al. (1991) found that 48 per cent of QTLs in tomato were detected in two environments. The results suggest that a large proportion of QTL affecting a quantitative trait in one environment will be active in other environments particularly for QTL with major effects. Mapping of such QTLs will make positive contribution to compare the phenotype in different environments. Tanksley (1993) has elaborated the procedures for high resolution mapping of QTLs. Fine mapping of QTL involves comparing the means of individual recombinants for molecular markers in the vicinity of the QTL with individuals that are non-recombinant.

The number of QTL varies depending upon the trait, nature of parents and the phenotypic variance. The small effect of QTLs can also be detected by the molecular marker method depending upon (i) the map distance from nearest marker to QTL; the closer a QTL is to a marker, the smaller the effects of QTL can be detected (ii) size of segregating population; the larger the population size, the more likely the effects of lesser QTL will reach statistical significance (iii) heritability; lower the heritability, the less likely a QTL will be detected. Tanksley (1993) indicated that in experiments with smaller population sizes and higher probability thresholds, QTLs that describe less than 3 per cent of the phenotypic variance are not normally detected.

Future research should focus on (1) identification of QTLs governing tolerance to major abiotic stresses, yield, yield components and quality traits particularly those having larger effects and which could be exploited in different environments for selection purposes (2) exploitation of complementary QTL to isolate transgressive segregants particularly from interspecific crosses (3) identification of orthologous QTL among different species; conservation of QTL for traits such as drought tolerance among species may provide new opportunities for manipulation of economic traits (4) high resolution QTL should help to determine whether QTL are single genes or clusters of tightly

linked genes and whether over dominance plays a significant role in heterosis (5) cloning of QTLs based on high resolution mapping will usher a new era in molecular quantitative genetics.

Molecular markers are particularly useful in fine mapping of QTL using overlapping chromosome substitution lines. Substitution mapping reduces the size of interval to which QTL can be assigned more precisely. Paterson et al. (1990) used this approach and mapped QTLs governing soluble solids concentration in intervals of as little as 3 cM by which the substituted segments differed. In this method, chromosomal segments in particular substitution line carrying the RFLP are identified and their regions of overlap determined using molecular markers. Phenotypic effects of each chromosomal segments are determined by QTL analysis of segregating progeny populations. Effects shared by different segments are attributed to QTLs in regions shared by those segments, while effects unique to a particular segments are attributed to QTLs in the region unique to that segment.

DNA MARKERS FOR IDENTIFICATION AND INTROGRESSION OF NOVEL QTLs

Molecular markers have been of great value to identify novel QTLs from unadapted and exotic germplasm that can enhance grain yield. Such QTLs have been introgressed into elite cultivars of tomato and rice. De Vicente and Tanksley (1993) observed transgressive segregation for 8 of 11 traits examined in an interspecific F₂ population derived from the cross of *Lycopersicon esculentum* x *L. pennelli*. RFLP analysis detected 74 QTL of which 30 per cent were related to the appearance of transgressive segregation. Tanksley and Nelson (1996) proposed advanced backcross QTL (AB-QTL) analysis method for simultaneous discovery and transfer of valuable QTLs from unadapted germplasm into elite breeding lines. AB-QTL analysis has been used to identify useful QTLs in rice (Xiao et al. 1998, Moncada et al. 2001). Xiao et al. (1996) analyzed BC₂ test cross families from the interspecific cross (*O. sativa* x *O. rufipogon*) and found that *O. rufipogon* alleles at marker loci RM5 on chromosome 1 and RM256 on chromosome 2 were associated with enhanced yield. The phenotypic advantage of the lines carrying *O. rufipogon* alleles (*yld1*,

yld2) at these loci, corresponds to an 18 per cent and 17 per cent increase over *O. sativa* parent. Xiao et al (1998) identified 68 QTLs in rice of which 35 had beneficial alleles derived from a phenotypically inferior *O. rufipogon* parent. Moncada et al. (2001) used advanced backcross QTL analysis to identify QTLs in BC₂F₂ populations of *Oryza sativa* x *O. rufipogon*. Based on analyses of 125 SSP and RFLP markers two putative *O. rufipogon* derived QTL were detected for yield, 13 for yield components and four for maturity. Furthermore, 56% of the trait enhancing QTL under upland environment were derived from *O. rufipogon*. Earlier 51% of favourable QTLs from this cross were identified under irrigated environments (Xiao et al. 1996). Molecular markers are of considerable interest to identify and introgress favourable QTLs from wild species into cultivated species, thus marker assisted breeding could be powerful in enhancing yield potential of crops. Comparative mapping among different cereals has also increased the efficiency of orthologous QTL mapping.

DNA MARKERS FOR MONITORING ALIEN INTROGRESSION AND DEVELOPMENT OF NOVEL GENETIC RESOURCES FOR GENE DISCOVERY AND FUNCTIONAL GENOMICS

DNA markers offer considerable advantages in backcross breeding (1) indirect selection of desirable gene(s) from donor parents, (2) selection for regions of concurrent genomes, and (3) selection of lines with minimum linkage drag near the introgressed regions. The target alien genes introduced into cultivated parents by backcross breeding are usually flanked by introgressed segments of DNA derived from the donor (exotic or wild) parent. This phenomenon (linkage drag) is frequently observed in alien introgression and affects traits other than the target genes. Through conventional backcrossing, it is difficult to reduce the size of the introgressed segments around the target gene. Molecular markers have unique advantage of overcoming the linkage drag. Young and Tanksley (1989) analyzed several lines where *Tm-2* gene which confers resistance to tobacco mosaic virus was introduced into cultivated tomato, *Lycopersicon esculentum* from *L. peruvianum* through backcrossing. RFLP analysis showed smallest introgressed segment to be 4 cM while the longest segment was 51 cM. RFLP markers can be effectively used to monitor the introgressed segments during backcrossing. It is estimated that by using molecular markers an

introgressed segment can be obtained in two generations that is as small as that which would otherwise take 100 generation without RFLP selection.

Tanksley and Hewitt (1988) analyzed tomato lines carrying segments introgressed from *L. chmielewskii*. These introgressed segments result into increased soluble solids in tomato varieties (Rice 1974). RFLP analysis showed that two of the three segments were associated with an increase in soluble solids. These segments were associated with deleterious effects showing increase in fruit pH, lower yield and smaller fruit. Alien segments introgressed from wild species have been identified in the advanced backcross progenies derived from crosses of *O. sativa* x *O. australiensis* and *O. sativa* x *O. brachyantha* (Ishii et al. 1994; Brar et al. 1996).

Bernachi et al. (1998) used MAS to select NILs containing a single introgressed segment from wild species into the genetic background of cultivated tomato. Monforte and Tanksley (2000) used molecular markers to analyze isogenic lines (NILs) and backcross recombinant inbred lines (BCRILs) derived from *Lycopersicon esculentum* x *L. hirsutum*. Most of the lines contained a single defined introgression from *L. hirsutum* in the *L. esculentum* background and together, the lines provided a coverage of more than 85% of the *hirsutum* genome. Such lines are important for genetic resources for gene discovery and functional genomics.

Doi et al. (1997) reported development of substitution lines with defined chromosome segments of *O. glaberrima* into the genetic background of *O. sativa*. RFLP markers were used to select plants with introgressed *glaberrima* segment during backcross generation. Lines were selected which were homozygous for the segments introgressed from *O. glaberrima*. Such substitution lines represent useful genetic resources for mapping of agronomically important genes and functional genomics studies. Chen and Slepper (1999) used FISH and RFLP markers to monitor introgression of *Festuca mairei* chromosomes into *Lolium perennis*.

Doganlar et al. (2000) identified RAPD marker (OPAB-06) linked to the reduced ripening time QTL on chromosome 12 of tomato. RAPD marker has

been converted to a cleaved amplified polymorphism (CAP) assay for MAS and introgression of reduced ripening time into commercial cultivars of tomato.

One of the major problems in using molecular markers in breeding programs is the expense involved. Protocols for large scale DNA extraction and reliable PCR reactions could make the technology more efficient as well as cost effective.

PRACTICAL CONSIDERATIONS IN MAS

The practical application of MAS depends upon several factors. Some of these are discussed below:

Tight linkage of the target trait(s) with molecular marker(s): One of the most important requirements to use MAS is the tight linkage between marker and target trait. This is important when segregating populations are being screened for the target traits through linked molecular markers. Such plants selected through MAS are continuously used in subsequent backcrossing program. This is easily achievable with major genes. However, it is still beyond practical use to apply MAS routinely for QTLs and complex agronomic traits of major interest to breeders.

Polymorphism: Polymorphism is essential to discriminate between the needed and unwanted individuals. Some markers show limited polymorphism particularly among closely related parents. However, microsatellite markers are by far the most polymorphic and can be used even among closely related parents or species.

Reproducibility: RAPD markers need well standardized conditions for reproducibility. However, other markers such as RFLP or SSR are easy to reproduce. It is expected that with the major efforts underway, the sequence of genomes of major crops such as rice, maize and wheat would soon become available. Thus, in future markers based on specific sequence differences, such as cleaved amplified polymorphism (CAPs) and single nucleotide polymorphism (SNPs), will become the markers of choice. Now, microsatellite markers are certainly the best choice for most purposes. These markers are highly

polymorphic, reliable, and abundantly available (McCouch et al. 2001). SSRs that can be scored using agarose gels are particularly useful in marker-assisted breeding.

Expense of MAS: The relatively high expense is another factor limiting the development and application of MAS. Standard procedures are multi-step and individually processed. These procedures are not suitable for MAS because of length of time required to grind the tissue, difficulty in handling several steps using multi-step procedures. The expense includes not only the materials and supplies but also less definable costs such as quality of technical support, lab space, and radioisotope permits. PCR-based markers such as microsatellites are amenable to automation. Recently, major improvements have been made in large scale extraction of DNA protocols. Lange et al. (1998) modified DNA isolation procedure that immobilizes DNA on a solid matrix followed by processing in a 96 well microtiter plate. A single person can isolate DNA and initiate PCR based analysis for 96 plants in one day. After rubbing a small amount of plant tissue on to a collection card, several small disks are punched off and the disk bound is partially purified in microtubes of 96 well microtiter plates. The discs can then be used directly in various PCR based DNA marker system such as RAPD, AFLP and microsatellites. This protocol may be modified for robotics also.

Number of genes in the screening program: The number of genes (loci) involved in MAS program is another factor that should be considered. Mackill et al. (1999) have elaborated the population size required based upon the number of genes involved in MAS. For example, with only four or five loci being selected, the population size and number of F₁ seeds needed for a MAS will be considerable and any further addition will lead to an exponential increase. The most important traits or loci should be identified and selected through MAS program.

CONCLUSIONS

Availability of high density molecular genetic maps covering entire genome and tight linkage of agronomic traits with PCR based markers which are fast, reliable and cost effective and automated procedures for DNA extraction have made

marker assisted breeding feasible. Furthermore, different kinds of mapping populations and genetic resources such as segregating populations, recombinant inbred lines (RIL), dihaploid (DH) lines, near isogenic lines (NILs), alien introgression lines with defined chromosome regions and overlapping chromosome substitution lines have greatly facilitated marker assisted breeding. Advances in molecular marker technologies have opened new era in gene discovery. MAS is becoming a reality for traits controlled by major genes. One of the important challenges is to apply MAS for complex agronomic traits such as yield, yield components and tolerance to abiotic stresses, which are highly influenced by the environments and are difficult to select through conventional plant breeding procedures. Much more efforts are needed by the molecular geneticists to precisely map such traits commonly referred as QTL before plant breeders could make use of these technologies. Integration of molecular marker assisted breeding and genomics with the on-going crop improvement programs is critical for indirect selection of desirable lines in different breeding programs which would be based on genotype rather than phenotype. Continued efforts are needed to make molecular marker assisted breeding cost effective so that it becomes a valuable addition to the existing tool box of plant breeders. Marker-assisted selection should be considered as good supplement to conventional selection procedures used in different breeding programs.

In future, the identification of the DNA sequences (genome sequence) and resulting proteins (proteomics) will further facilitate MAS for the target traits. Once gene sequences become known, it will lead to the use of more precise markers such as SNPs in marker-assisted breeding.

REFERENCES

- Bernacchi, D., T. Beck-Bunn, D. Emmatty, Y. Eshed, S. Inai, J. Lopez, V. Petiard, H. Sayama, J. Uhlig, D. Zamir and S.D. Tanksley. 1998. Advanced backcross QTL analysis of tomato. II. evaluation of near-isogenic lines carrying single-donor introgressions for desirable wild QTL-alleles derived from *Lycopersicon hirsutum* and *L. pimpinellifolium*. *Theor Appl Genet.* 97:170-180.
- Bernatzky, R and S.D. Tanksley. 1986. Toward a saturated linkage map of tomato based on isozyme and random cDNA sequences. *Genetics.* 122:887-898.

- Brar, D.S. and H.S. Dhaliwal. 1997. Molecular markers and their application in crop improvement (eds. M.S. Bajwa, J.S. Dhillon, V.K. Dilawari and S.S. Chahal). *Proc. Third Agricultural Science Congress. Invited Papers, Vol. I*, National Academy of Agricultural Sciences, New Delhi, India, pp. 175-192.
- Brar, D.S., R. Dalmacao, R. Elloran, R. Aggarwal, R. Angeles, and G.S. Khush. 1996. Gene transfer and molecular characterization of introgression from wild *Oryza* species into rice. In: *Rice Genetics III*. Pp. 47-86. International Rice Research Institute, Manila, Philippines.
- Chen, C. and D.A. Sleper. 1999. FIH and RFLP marker-assisted introgression of *Festuca mairei* chromosomes into *Lolium perenne*. *Crop Sci.* 39:1676-1679.
- Concibido, V.C., R.L. Denny, D.A. Lange J.H. Orf and N.D. Young. 1996. RFLP mapping and marker-assisted selection of soybean cyst nematode resistance in PI 209332. *Crop Sci.* 36:1643-1650.
- De Vicente, M.C. and S.D. Tanksley. 1993. QTL analysis of transgressive segregation in an interspecific tomato cross. *Genetics*. 134:585-596.
- Doganlar, S., S.D. Tanksley and M.A. Mutschler. 2000. Identification and molecular mapping of loci controlling fruit ripening time in tomato. *Theor Appl Genet.* 100:249-255.
- Doi, K., N. Iwata and A. Yoshimura. 1997. The construction of chromosome substitution lines of African rice (*Oryza glaberrima* Steud) in the background of japonica rice (*O. sativa* L.). *Rice Genet Newsl.* 14:39-41.
- Dweikat, I., H. Ohm, F. Patterson and S. Cambron. 1997. Identification of RAPD markers for 11 Hessian fly resistance genes in wheat. *Theor Appl Genet.* 94:419-423.
- Frisch, M., M. Bohn and A.E. Melchinger. 1999a. Minimum sample size and optimal positioning of flanking markers in marker-assisted backcrossing for transfer of a target gene. *Crop Sci.* 39:967-975.
- Frisch, M., M. Bohn and A.E. Melchinger. 1999b. Comparison of selection strategies for marker-assisted backcrossing of a gene. *Crop Sci.* 39:1295-1301.
- Harushima, Y., M. Yano, A. Shomura, M. Sato, T. Shimano, Y. Kuboki, T. Yamamoto, S.Y. Lin, B.A. Antonio, A. Parco, H. Kajiya, N. Huang, K. Yamamoto, Y. Nagamura, N. Kurata, G.S. Khush and T. Sasaki. 1998. A high-density rice genetic linkage map with 2275 markers using a single F₂ population. *Genetics*. 148:479-494.
- Hittalmani, S., A. Parco, T.V. Mew, R.S. Zeigler and N. Huang. 2000. Fine mapping and DNA marker-assisted pyramiding of the three major genes for blast resistance in rice. *Theor Appl Genet.* 100:1121-1128.
- Huang, N., E.R. Angeles, J. Domingo, G. Magpantay, S. Singh, G. Zhang, N. Kumaravadivel, J. Bennett and G.S. Khush. 1997. Pyramiding of bacterial

- blight resistance genes in rice: marker-assisted selection using RFLP and PCR. *Theor Appl Genet.* 95:313-320.
- Ishii, T., D.S. Brar, D.S. Multani, and G.S. Khush. 1994. Molecular tagging of genes for brown planthopper resistance and earliness introgressed from *Oryza australiensis* into cultivated rice, *O. sativa*. *Genome.* 37:217-221.
- Katiyar, S.K., Y. Tan, B. Huang, G. Chandel, Y. Xu, Y. Zhang, Z. Xie and J. Bennett. 2001. Molecular mapping of gene *Gm6(t)* which confers resistance against four biotypes of Asian rice gall midge in China. *Theor Appl Genet.* (in press).
- Khush, G.S. and D.S. Brar. 1998. The application of biotechnology to rice. In: *Agricultural Biotechnology in International Development*. pp. 97-121. (eds. C.L. Ives and B.M. Bedford). CAB International, Wallingford, UK.
- Lange, D.A., S. Peñuela, R.L. Denny, J. Mudge, V.C. Concibido, J.H. Orf and N.D. Young. 1998. A plant DNA isolation protocol suitable for polymerase chain reaction based marker-assisted breeding. *Crop Sci.* 38:217-220.
- Leblanc, O., D. Grimanelli, D. Gonzalez de Leon and Y. Savidan. 1995. Detection of the apomictic mode of reproduction in maize-*Tripsacum* hybrids using maize RFLP markers. *Theor Appl Genet.* 90:1198-1203.
- Lee, M. 1995. DNA markers and plant breeding programs. *Adv Agron.* 55:265-344.
- Leung, H.R., J. Nelson and J.E. Leach. 1993. Population structure of the plant pathogens fungi and bacteria. *Adv Plant Path.* 10:158-205.
- Li, Z. 2001. QTL mapping in rice: a few critical considerations. In: *Proc. 4th Intern Rice Genet Symp.* (In-press).
- Mackill, D. J., H. T. Nguyen and J. X. Zhang. 1999. Use of molecular markers in plant improvement programs for rainfed lowland rice. *Field Crops Res.* 64:177-185.
- Mackill, D.J. and J. Ni. 2001. Molecular mapping and marker-assisted selection for major gene traits in rice. In: *Proc. 4th Intern Rice Genet Symp.* (In-press).
- McCouch, S. R., S. Temnykh, A. Lukashova, J. Coburn, G. Declerck, S. Cartinhour, S. Harrington, M. Thompson, E. Septiningsih, M. Semon, P. Moncada, J. Li and E. Paul. 2001. Microsatellites in rice: abundance, diversity and applications. In: *Proc. 4th Intern Rice Genet Symp.* (In-press).
- Michelmore, R.W., I. Paran and R.V. Kesseli. 1991. Identification of markers linked to disease resistance genes by bulked segregant analysis: a rapid method to detect markers in specific genomic regions by using segregating populations. *Proc Natl Acad Sci USA.* 88:9828-9832.
- Miklas, P.N., R.C. Larsen, R. Riley and J.D. Kelly. 2000. Potential marker assisted selection for *bc-1²* resistance to bean common mosaic poty virus in common bean. *Euphytica* 116:211-219.

- Mohan, M., S. Nair, A. Bhagwat, T.G. Krishna, M. Yano, C.R. Bhatia and T. Sasaki. 1997. Genome mapping, molecular markers and marker-assisted selection in crop plants. *Mol Breed.* 3:87-103.
- Moncada, P., C.P. Martinez, J. Borrero, M. Chatel, H. Gauch, E. Guimaraes, J. Tohme and S.R. McCouch. 2001. Quantitative trait loci for yield and yield components in an *Oryza sativa* x *O. rufipogon* BC₂F₂ population evaluated in an upland environment. *Theor Appl Genet.* 102:41-52.
- Monforte, A.J. and S.D. Tanksley. 2000. Development of a set of near isogenic and backcross recombinant inbred lines containing most of the *Lycopersicon hirsutum* genome in a *L. esculentum* genetic background: A tool for gene mapping and gene discovery. *Genome.* 43:803-813.
- Ozias-Akins, P., E. L. Lubbers, W. W. Hanna and J. W. McNay. 1993. Transmission of the apomictic mode of reproduction in *Pennisetum*: co-inheritance of the trait and molecular markers. *Theor Appl Genet.* 85:632-638.
- Paterson, A. H. 1998. QTL mapping in DNA marker assisted plant and animal improvement. In: *Molecular dissection of complex traits.* pp. 131-143. (eds. A. H. Paterson). CRC Press, Boca Raton, USA.
- Paterson, A.H., S.D. Tanksley and M.E. Sorrell. 1994. DNA markers in plant improvement. *Adv Agron.* 46: 39-90.
- Paterson, A. H., S. Damon, J. D. Hewitt, D. Zamir, H. D. Rabinowitch, S. E. Lincoln, E. S. Landers and S. D. Tanksley. 1991. Mendelian factors underlying quantitative traits in tomato: comparison across species, generations and environments. *Genetics.* 127:181-197.
- Paterson, A.H., J.W. De Verna, B. Lanini and S.D. Tanksley. 1990. Fine mapping of quantitative trait loci using selected overlapping recombinant chromosomes in an interspecific cross of tomato. *Genetics.* 124:735-742.
- Paterson, A.H., E.S. Lander, J.D. Hewitt, S. Peterson, S.E. Lincoln and S.D. Tanksley. 1988. Resolution of quantitative traits into Mendelian factors, using a complete linkage map of restriction fragment length polymorphisms. *Nature.* 335:721-726.
- Penner, G.A., J.A. Stebbing and B. Legge. 1995. Conversion of an RFLP marker for the barley stem rust resistance gene *Rpg1* to a specific PCR-amplifiable polymorphism. *Mol Breed.* 1:349-354.
- Ribaut, J.M. and J. Betran. 1999. Single large-scale marker-assisted selection (SLS-MAS). *Mol Breed.* 5:535-446.
- Rick, C. M. 1974. High soluble-solids content in large-fruited tomato lines derived from a wild green fruited species. *Hilgardia.* 42:493-498.

- Rouf Mian, M.A., T. Wang, D.V. Phillips, J. Alvernaz and H.R. Boerma. 1999. Molecular mapping of the *Rcs3* gene for resistance to frog-eye leaf spot in soybean. *Crop Sci.* 39:1687-1691.
- Sanchez, A.C., D.S. Brar, N. Huang, Z. Li and G.S. Khush. 2000. Sequence tagged site marker-assisted selection for three bacterial blight resistance genes in rice. *Crop Sci.* 40:792-797.
- Sax, K. 1923. The association of size differences with seed-coat pattern and pigmentation in *Phaseolus vulgaris*. *Genetics.* 8:552-560.
- Singh, S., J.S. Sidhu, N. Huang, Y. Vikal, Z. Li, D.S. Brar, H.S. Dhaliwal and G.S. Khush. 2001. Pyramiding three bacterial blight resistance genes (*xa5*, *xa13* and *Xa21*) using marker-assisted selection into indica rice cultivar PR106. *Theor Appl Genet.* 102:1011-1015.
- Stubber, C.W., S.E. Loncoln, D.W. Wolff, T. Helentjaris, E.S. Lander. 1992. Identification of genetic factors contributing to heterosis in a hybrid from two elite maize inbred lines using molecular markers. *Genetics.* 132:823-839.
- Tanksley, S.D. 1983. Molecular markers in plant breeding. *Plant Mol Biol Rep.* 1:3-13.
- Tanksley, S.D. 1993. Mapping polygenes. *Annu Rev Genet.* 27:205-233.
- Tanksley, S.D. and J.D. Hewitt. 1988. Use of molecular markers in breeding for soluble solids in tomato - a re-examination. *Theor Appl Genet.* 75:811-823.
- Tanksley, S.D. and J.C. Nelson. 1996. Advanced backcross QTL analysis: a method for simultaneous discovery and transfer of valuable QTL from unadapted germplasm into elite breeding lines. *Theor Appl Genet.* 92:191-203.
- Tanksley, S.D., N.D. Young, A.H. Paterson and M.W. Bonierbale. 1989. RFLP mapping in plant breeding: new tools for an old science. *Bio/Technology.* 7:257-264.
- Thoday, J.M. 1961. Location of polygenes. *Nature.* 191:368-370.
- Thomas, C.M., P. Vos, M. Zabeau, D.A. Jones, K.A. Norcott, B.P. Chadwick and J.D.G. Jones. 1995. Identification of amplified restriction fragment length polymorphism (AFLP) markers tightly linked to the tomato, *Cf-9* gene for resistance to *Cladosporium fulvum*. *Plant J.* 8:785-794.
- Wang, G., D.J. Mackill, J.M. Bonman, S.R. McCouch, M.C. Champoux et al. 1994. RFLP mapping of genes conferring complete and partial resistance to blast in a durably resistant rice cultivar. *Genetics.* 136:1421-1434.
- Xiao, J., S. Grandillo, S.N. Ahn, S.R. McCouch, S.D. Tanksley, J. Li and L. Yuang. 1996. Genes from wild rice improve yield. *Nature.* 384:223-224.
- Xiao, J., J. Li, S. Grandillo, S.N. Ahn, L. Yuan, S.D. Tanksley and S.R. McCouch. 1998. Identification of trait-improving quantitative trait loci alleles from a wild rice relative, *Oryza rufipogon*. *Genetics.* 150:899-909.

- Xu, Y. 1997. Quantitative trait loci: separating, pyramiding and cloning. *Plant Breeding Reviews*. 15:1-34.
- Yenco, G.C., M.B. Cohen and P.F. Byrne. 2000. Applications of tagging and mapping insect resistance loci in plants. *Annu Rev Entomol*. 45:393-422.
- Yoshimura, S., A. Yoshimura, N. Iwata, S.R. McCouch, M.L. Abenes, M.R. Baraoidan, T.W. Mew and R.J. Nelson. 1995. Tagging and combining bacterial blight resistance genes in rice using RAPD and RFLP markers. *Mol Breed*. 1:375-387.
- Young, N.D. 1999. A cautiously optimistic vision for marker-assisted breeding. *Mol Breed*. 5:505-510.
- Young, N.D. and S.D. Tanksley. 1989. RFLP analysis of the size of chromosomal segments retained around the *Tm-2* locus of tomato during backcross breeding. *Theor Appl Genet*. 77:95-101.
- Zheng, K., N. Huang, J. Bennett and G.S. Khush. 1995. PCR-based marker assisted selection in rice breeding. *IRRI Discussion Paper Series* No. 12, International Rice Research Institute, Manila, The Philippines, pp. 24.

4

MICROSATELLITES AND MOLECULAR BREEDING: EXPLOITATION OF MICROSATELLITE VARIABILITY FOR THE ANALYSIS OF A MONOTONOUS GENOME

Peter Winter, Bruno Hüttel, Kurt Weising and Günter Kahl

Plant Molecular Biology, Department of Biology, Biocentre, University of Frankfurt am Main, D-60439 Frankfurt, Germany

INTRODUCTION

Molecular markers and especially those based on selectively neutral DNA sequence polymorphisms have revolutionized the old art of plant breeding. They facilitate the reliable identification of clones, breeding lines, hybrids and cultivars, allow the monitoring of introgression of alien DNA into cultivated germplasm, and the estimation of genetic diversity in germplasm collections. Further, for nearly every important crop, advanced high-density DNA marker maps are now available that provide a basis for marker-assisted selection of agronomically useful traits, pyramiding of resistance genes, and the isolation of these and other important genes by map-based cloning. In future, molecular marker techniques clearly will gain more and more influence on plant breeding, and will speed up breeding processes considerably (see reviews by Tanksley et al. 1995, Winter and Kahl 1995).

Not all types of DNA markers work equally well for every crop due to differences in breeding history, reproduction, and genome organization. For example, highly sensitive marker systems are required for the analysis of species or populations that are either obligatory self-pollinating, or went through genetic bottlenecks during domestication, or both. Species with such properties are often characterized by largely invariable genomes as has been shown e.g. for the common bean (Sonnante et al. 1994) and the cultivated tomato (Miller and Tanksley 1990). Since finding genetic polymorphisms in

such "monotonous" genomes is notoriously difficult, the generation of molecular linkage maps may be delayed tremendously.

In the present review, we would like to focus on chickpea (*Cicer arietinum* L.) as a model system for leguminous crop plants with a monotonous genome. Chickpea is the third most important grain legume crop worldwide, and is nowadays cultivated on the Indian sub-continent, in West Asia, North Africa, the Americas and Australia. Early molecular marker studies based on isozymes, seed protein patterns and/or restriction fragment length polymorphisms (RFLPs, Botstein et al. 1980) showed that the genetic spectrum among chickpea landraces and cultivars is narrow (Kazan and Muehlbauer 1991, Ahmad and Slinkard 1992; Udupa et al. 1993; Labdi et al. 1996). Two reasons account for this. First, the crop is a self-pollinated diploid ($2n=16$) showing high levels of inbreeding. Second, cultivated chickpea, most probably passed a genetic bottleneck upon selection from its presumable ancestor *C. reticulatum* by the first farmers around 10.000 years ago (van der Maesen 1987, Zohary and Hopf 1993). As a consequence of the low level of intraspecific variation, initial chickpea maps comprising a few isozyme, morphological, RFLP and random amplified polymorphic DNA (RAPD, Williams et al. 1990) markers are based on *interspecific* crosses between the cultigen and *C. reticulatum* (Gaur and Slinkard 1990, Kazan et al. 1993, Simon and Muehlbauer 1997). These maps were of relatively low density.

Notwithstanding the overall monotony of the chickpea genome, Weising et al. (1989, 1991, 1992) demonstrated considerable variation at the intra- and inter-species level by using simple repetitive DNA oligonucleotides as hybridisation probes in RFLP-type experiments. The target sequences of these experiments occur ubiquitously in eukaryotic genomes and are nowadays called **microsatellites** (MS, Litt and Luty 1989), **simple sequence repeats** (SSRs), or **short tandem repeats** (STRs). SSRs consist of short, reiterated sequence motifs of about 1 to 5 nucleotides that, like classical satellite DNA, are organised in more or less perfect tandem arrays of few to hundreds or even thousands of repeat units. A key feature of SSR repeats is a strong tendency to change their overall length by slipped-strand mispairing and other less well understood processes (Levinson and Gutman 1987, Zischler et al. 1992, for review see McMurray 1995, Sia et al. 1997) leading to **variable numbers of tandem repeats** (VNTR, Nakamura et al. 1987) and resulting in **simple sequence length polymorphisms** (SSLPs, Cho et al. 2000).

Mutation rates of SSRs are usually high but may vary considerably among loci and organisms. For example, 2.5×10^{-5} to 1×10^{-2} mutations per locus per gamete per generation have been estimated in mammalian genomes (Weber and Wong 1993) as compared to a combined average mutation rate of 6.5×10^{-6} in *Drosophila*, as calculated from 10 di-, 6 tri- and 8 tetranucleotide repeats (Schug et al. 1998). However, values exceeding the

average by two orders of magnitude were detected at certain SSR loci, suggesting locus-specific mutation rates in *Drosophila* (Schlötterer et al. 1998; Harr et al. 1998). Locus-, motif- or even allele-specific mutation rates were also demonstrated in humans (Jin et al. 1996, Chakraborty et al. 1997). To our knowledge, no mutation rates of nuclear SSRs have yet been published for any plant species, but the large number of alleles at certain loci (up to > 30; Saghai-Maroo et al. 1994; Weising et al. 1996) indicate that maximum mutation rates are similar in plants and animals.

SSRs are abundant and usually more or less evenly dispersed throughout the plant genome (for exceptions see Areshchenkova and Ganai 1999) but appear to be less frequent in plants as compared to vertebrates (Lagercrantz et al. 1993). Reported estimates of microsatellite frequencies vary considerably among plant species, with average inter-SSR distances ranging from 10 kb to more than 1 Mb, depending on the motif and the organism (e.g., Morgante and Olivieri 1993, Bell and Ecker 1994, Becker and Heun 1995a; Panaud et al. 1995; Broun and Tanksley 1996, Pfeiffer et al. 1997). Significant differences in SSR abundance were observed between algae, monocots and dicots (Wang et al. 1994). Thus, SSRs with a minimum length of 20 bp were found every 19 kb on average in algae, every 65 kb in monocots, and every 21 kb in dicots. In the chickpea genome $[TAA]_n$, $[GA]_n$ and $[CA]_n$ arrays are present at more than 12,000 loci with an average spacing of around 60 kb (Hüttel et al. 1999).

SSRs are generally rare in chloroplast and mitochondrial genomes. Only four SSRs exceeding a size of 20 bp were found in an early database search, all of which were located in chloroplast DNA (Wang et al. 1994). More recently, SSRs have also been identified in the mitochondrial genome of several gymnosperms (Soranzo et al. 1999). Most organellar SSRs are short- to medium-sized stretches of poly[A] or poly[G] (Powell et al. 1995a,b; Weising and Gardner 1999; Soranzo et al. 1999). Depending on their location, chloroplast SSRs are variable at different taxonomic levels (Bryan et al. 1999; Weising and Gardner 1999; Ishii and McCouch 2000), and markers based on chloroplast SSRs have been used for population genetic studies in various species, mainly gymnosperms (e.g., Echt et al. 1998; Vendramin et al. 1998). Despite their lower mutation rates as compared to most nuclear SSRs (estimated to be 3.2 to 7.9×10^{-5} in a highly monomorphic species such as *Pinus torreyana*; Provan et al. 1999a), the usefulness of chloroplast SSRs for taxonomic studies at the species level is probably limited by homoplasy (Doyle et al. 1998), as is the case with nuclear microsatellites (Peakall et al. 1998).

While most SSRs are probably non-functional and, therefore, selectively neutral (Tachida and Iizuka 1992, Epplen et al. 1998, Stephan and Kim 1998), numerous studies have suggested possible roles in gene regulation, chromatin architecture, recombination, replication and/or evolutionary adaptation (e.g., Hamada et al. 1984, Kashi et al. 1997, Meloni

et al. 1998, Mäueler et al. 1999). For example, studies of microsatellite loci in wild *Triticum dicoccoides* populations in neighbouring microhabitats in Israel revealed a non-random, habitat-specific distribution of alleles suggesting a role in adaptation to local environments (Li et al. 2000). High levels of polymorphism, abundance and ubiquitous occurrence recommend SSRs as exceptionally useful molecular markers (see reviews by Weising et al. 1995a, 1998; Powell et al. 1996, Gupta and Varshney 2000). Using chickpea as an example, we will portray the different SSR-based marker methodologies currently available and discuss their applications .

1. MICROSATELLITE-BASED TECHNIQUES TO EXPLOIT A RICH AND EASILY ACCESSIBLE SOURCE OF GENETIC POLYMORPHISM

Techniques for the detection of SSR polymorphism can basically be classified into two main categories, i.e. (1) hybridisation-based methods and (2) PCR-based methods. **Hybridization-based methods** make use of (non)radioactively labelled, synthetic SSR-specific oligonucleotide probes for multilocus RFLP fingerprinting or hybridisation to electrophoretically resolved RAPD fragments. The various **PCR-based methods** differ mainly in the sequences and position of primers relative to the SSR motifs as outlined in Fig.1. In most of these methods, SSR-complementary oligonucleotides serve as PCR primers, either alone or in combination with arbitrary or specific primers, to amplify certain regions of genomic DNA.

Methods based on SSR-complementary primers generally result in dominant markers that can only rarely be transferred from one population to the other, and moreover, are unable to reliably detect heterozygotes. The only SSR-based technique that routinely generates codominant markers is the locus-specific amplification of SSRs from specifically designed primers directed towards their flanking sequences. With the importance of the latter type of markers in mind, we will discuss these in some detail.

1.1. Multilocus RFLP Fingerprinting

The earliest approach to exploit SSR variability was developed by Epplen and colleagues (Ali et al. 1986). It relies on hybridisation of end-labelled oligonucleotide probes complementary to SSR motifs (e.g. [GATA]₄) to enzymatically restricted and electrophoretically separated genomic DNA. Hybridization generates highly informative, multiple banding patterns, so-called multilocus RFLP fingerprints, that often allow the distinction of individuals. The use of oligonucleotide probes is compatible

with in-gel hybridisation which is faster and more efficient than Southern blotting (see Epplen 1992, Weising and Kahl 1997 for a detailed description of the methodology). RFLP analysis with SSR-specific oligonucleotides is also called "oligonucleotide fingerprinting"

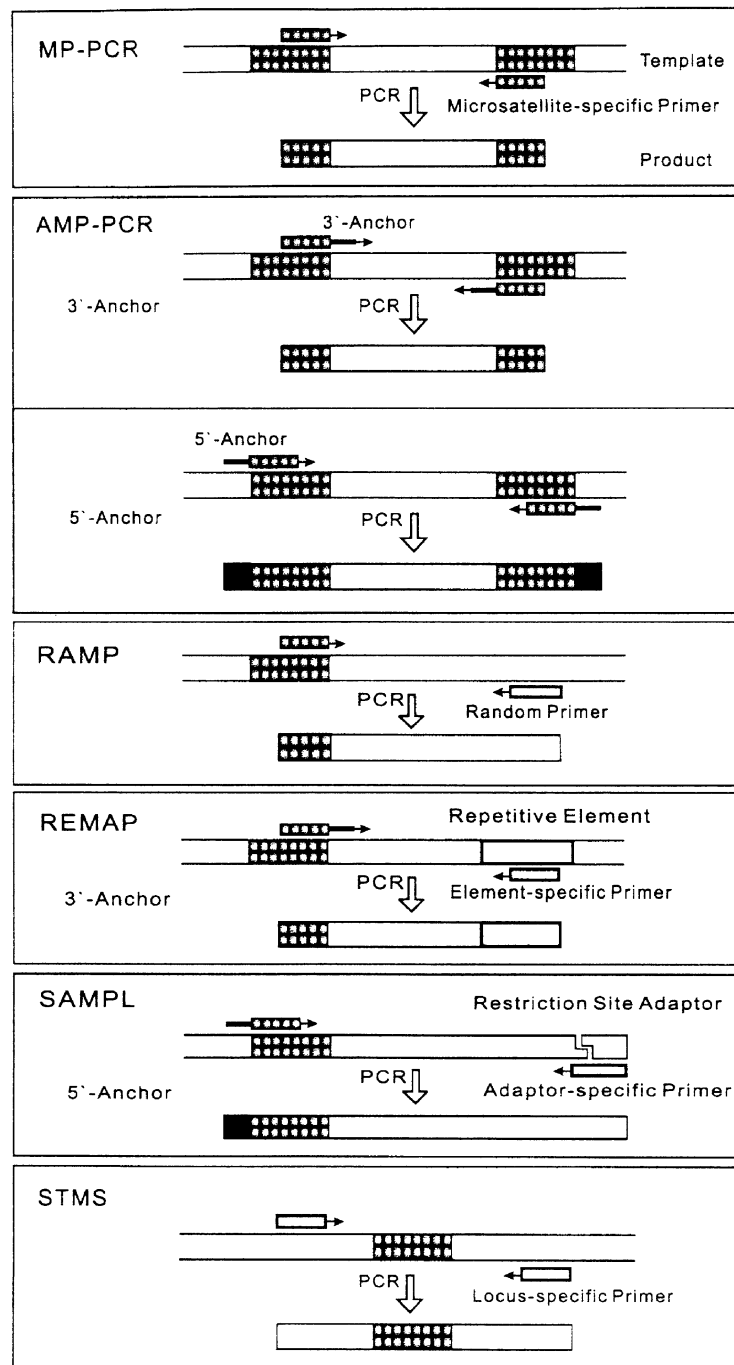


Figure 1: A schematic survey of PCR-based SSR marker methodologies. The different techniques mainly differ in the position and type of primer(s) used for amplification. See text for details.

The probes used in RFLP fingerprinting certainly recognize SSR-like target sequences. However, the molecular basis of the observed RFLP-type polymorphisms is not yet perfectly clear. While sequenced SSRs generally span about 20 to 500 bp, polymorphic RFLP fingerprint bands are often much larger (up to more than 10 kb), and their size may vary by several kilobases. Thus, the observed polymorphisms are most likely caused by the cumulative action of VNTR mutations, restriction site variation, and possibly transposition events (for a discussion see e.g. Zischler et al. 1992, Broun and Tanksley 1993, Kashi et al. 1994, Vosman and Arens 1997).

RFLP fingerprinting has been applied to a large number of plant, fungal and animal species and for many different purposes (reviewed by Weising et al. 1995a). For example, sex-specific polymorphisms have been detected by fingerprinting especially with the [GATA] motif (e.g. in guppy fish: Nanda et al. 1990; or papaya: Parasnis et al. 1999). The effects of different combinations of probe, enzyme and target DNA have been most comprehensively studied in *Arabidopsis thaliana* (Depeiges et al. 1995) and in chickpea (Weising et al. 1992; Sharma et al. 1995a,b). It should be noted, that not all nucleotide combinations are useful. For example, of the 353 theoretically possible di-, tri-, and tetranucleotides only about one tenth is practically applicable, since (1) many probes are redundant (e.g. each of the probes [GAT]₅, [ATG]₅, [TGA]₅, [ATC]₅, [TCA]₅ and [CAT]₅ will detect either one or the other strand of a locus consisting of six or more GAT-repeats), and (2) some motifs such as [GGCC]_n, [AT]_n or [GC]_n are self-annealing.

In chickpea, four accessions of the crop were analysed with 14 different restriction enzymes and 38 di, tri and tetranucleotide probes. Of these, 35 yielded detectable hybridization signals. The abundance and polymorphism of the target sequences varied considerably (see Fig. 2). No obvious correlation existed between abundance, fingerprint quality, and sequence characteristics of a particular motif (Sharma et al. 1995a). On the basis of hybridization patterns the probes were classified into 3 categories:

- Category I probes (e.g. [GATA]₄, [GCGT]₄, [GAAT]₄, or [GTGA]₄) yielded distinct, polymorphic banding patterns of probe-specific complexity.
- Category II probes (as e.g. [GATT]₄) yielded distinct, monomorphic banding patterns.
- Category III probes ([GTAA]₄) yielded either a smear, diffuse bands superimposed on a high in-lane background, or very weak signals only.

The fragment numbers strongly depended on the probe, less on the enzyme. The category I probes could be further subdivided according to the detected level of variability (see Fig. 2). Thus, some probes like $[GATA]_4$ or $[CAA]_5$ revealed considerable between- as well as intra-accessional polymorphism. Others, such as $[GCGT]_4$, differentiated between accessions, but produced monomorphic patterns within an accession. SSR variation is generally higher between than within accessions. The complete set of twenty category I probes detected 78 screenable *TaqI*-polymorphisms between two accessions. Digestion with methyl-sensitive enzymes revealed that simple sequence motifs are enriched in highly methylated genomic regions.

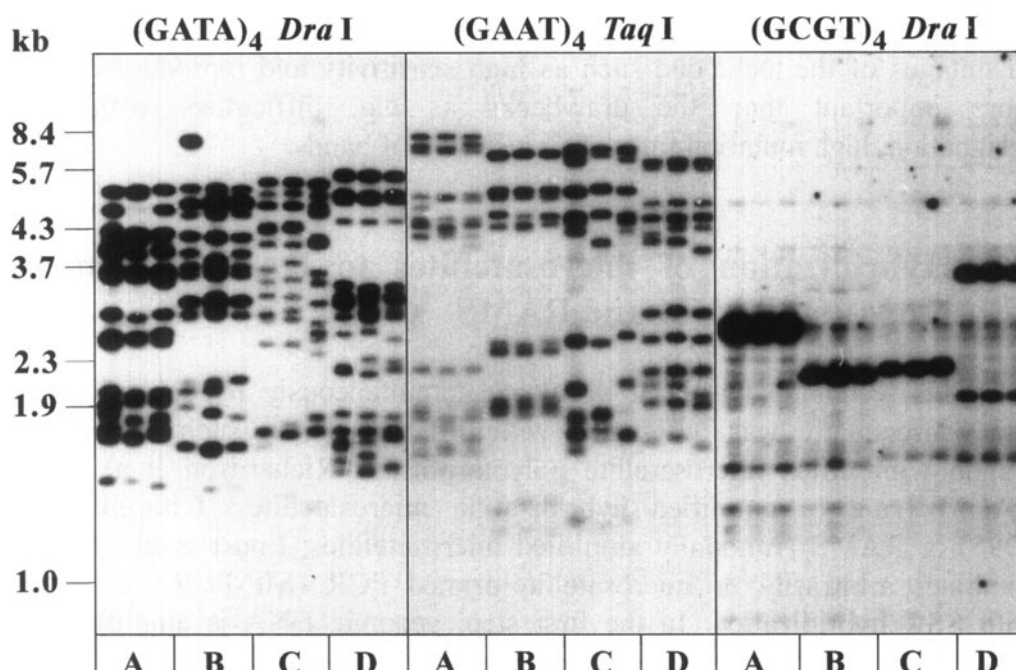


Figure 2: Oligonucleotide fingerprinting of chickpea accessions. DNA was isolated from three randomly selected individual plants of each of four accessions (A: ILC 482; B: ILC 1272; C: ILC 1929; D: ILC 3279), digested with *Taq I* or *Dra I* and separated in 1.2% agarose gels (5 ug DNA per lane). DNA in the gels was then denatured, neutralized, the gels dried, and consecutively hybridized to several SSR-complementary oligonucleotide probes. The results obtained with $[GATA]_4$, $[GAAT]_4$ and $[GCGT]_4$ are shown here. Note that $[GCGT]_4$ produces accession-specific patterns, while intra-accessional variability becomes apparent with $[GATA]_4$ and $[GAAT]_4$. Positions of molecular weight markers are given in kilobases.

The mitotic stability of RFLP fingerprints made the technique highly useful for the unequivocal identification and discrimination of vegetatively propagated plant material such as banana cultivars (Kaemmer et al. 1992), and micropropagated yarrow (*Achillea*) plants (Wallner et al. 1996). Basic problems are, however, encountered when fingerprint data are applied to

genetic mapping. One serious problem concerns mutation rate. Considerable proportions of nonparental fingerprint bands indicative of extremely high mutation rates (change in fragment size from one generation to the next) were observed in RFLP fingerprints of several plant species, including chickpea (Hüttel 1996, Winter, unpublished). A second problem for mapping is a distinct tendency of long SSR arrays (that are presumably detected by RFLP fingerprinting) to occur in clusters, unlike the much shorter SSRs usually analysed by PCR (see below). Such clustering was observed in pea (Dirlewanger et al. 1994) and tomato, where large SSRs appear to be associated with centromeric regions (Arens et al. 1995; Broun and Tanksley 1996, Areshchenkova and Ganai 1999).

Keeping these problems in mind, the identification and discrimination of genotypes (e.g. cultivars and clones) is probably the most appropriate application for RFLP fingerprinting in plant breeding, because here the advantages of the technique such as high sensitivity and reproducibility are more important than the drawbacks as e.g. difficulties with allele designation, high mutation rates and clustering of bands.

1.2. Hybridization of microsatellites to RAPD fragments: RAMPO, RAHM and RAMS

This technique has been developed simultaneously by several groups and consequently was designated with different names such as RAMPO (**r**andom **a**mplified **m**icrosatellite **p**olymorphisms; Richardson et al. 1995), RAHM (**r**andom **a**mplified **h**ybridization **m**icrosatellites; Cifarelli et al. 1995) or RAMS (**r**andomly **a**mplified **m**icrosatellites; Ender et al. 1996). It combines arbitrarily or **m**icrosatellite-primed PCR (MP-PCR, see below) with SSR hybridization. In the first step, genomic DNA is amplified with either a short primer of random sequence (as in RAPD analysis; Williams et al. 1990), or an SSR-complementary primer (Gupta et al. 1994; Weising et al. 1995b). PCR products are then separated on agarose gels, blotted onto membranes, and hybridized to ³²P- or digoxigenin-labelled mono-, di-, tri- or tetranucleotide repeat probes. The resulting banding patterns are reproducible, probe-dependent fingerprints which are often completely different from the ethidium bromide stained bands, and are polymorphic at an intraspecific level (Ramser et al. 1997).

RAMPO bands are probably derived from the many minor RAPD or MP-PCR reaction products that are usually hidden below the ethidium bromide detection level. Due to the high abundance and ubiquitous presence of SSRs in eukaryotic genomes, sensitive hybridisation with microsatellite motifs visualizes a subset of such minor amplification products. Hybridising bands may be isolated, cloned, sequenced and either used as a probe on

RAPD or RFLP gels (Cifarelli et al. 1995), or flanking sequences exploited for the generation of primers for locus-specific SSR PCR (Ender et al. 1996).

1.3. Microsatellite-primed PCR with unanchored primers: MP-PCR and SPAR

Oligonucleotides containing SSR motifs are not only used as hybridization probes (as described above), but may also serve as single PCR primers. These primers anneal at the 3'-ends of microsatellite loci and allow the amplification of the inter-simple sequence repeat (ISSR) regions, if inversely oriented SSRs are located on opposite strands close enough for amplification (see Fig. 1). These ISSRs are separated on agarose gels, and stained with ethidium bromide. Polymorphisms detected that way are not due to microsatellite length variation in most cases, but are often caused by small insertions/deletions in intervening sequences. The method was initially developed by Meyer et al. (1993) for the differentiation of strains and serotypes of the human fungal pathogen *Cryptococcus neoformans*. Later on, it was also applied to other fungi, as well as animals and plants.

Using an MP-PCR variant called SPAR (single primer amplification reaction), Gupta et al. (1994) tested 23 primers complementary to di-, tri-, tetra- and pentanucleotide repeats for their ability to amplify genomic DNA across a panel of eukaryotes. There, tetranucleotide repeat primers were most efficient in amplifying polymorphic patterns. GC- as well AT-rich primers worked equally well. Primers representing a combination of two tetranucleotide repeats, or compound SSRs were also effective. Single base permutations yielded different PCR fingerprints. In genetically relatively invariable popcorn (*Zea mays* L.) hybrids, 54 ISSR bands were scored per primer per lane, of which up to 87 % were polymorphic (Kantety et al. 1995).

In our own studies using a variety of di-, tri-, and tetranucleotide repeat-containing oligonucleotides as PCR primers for the analysis of yeast, human and different plant DNAs (Weising et al. 1995b, Sharma et al. 1995b), distinct, polymorphic banding patterns were obtained with most primers in all investigated species including chickpea (Fig. 4). From 1 to about 20 bands were obtained, mostly in the range between 0.3 and 2 kb. However, in plant DNAs dinucleotide repeats as well as AT-rich trinucleotide repeats often produced a smear, probably a consequence of the large copy number of these motifs in plants (see below). In these experiments the reproducibility and sensitivity to reaction conditions of MP-PCR was similar to RAPD analysis, since the majority of MP-PCR bands originated from mismatch annealing of primers as is also expected for RAPD bands (Weising et al. 1995b).

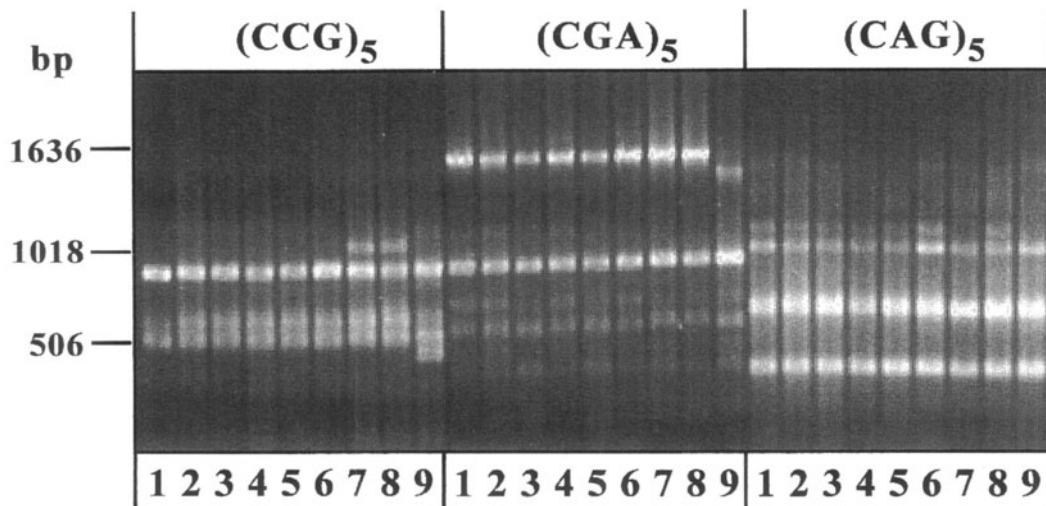


Figure 3: MP-PCR analysis of intra- and interspecific variation in chickpea. Genomic DNA from individual plants was amplified using the unanchored trinucleotide repeats $[CCG]_5$, $[CGA]_5$ and $[CAG]_5$ as single primers. PCR products were separated on 1.5% agarose gels and stained with ethidium bromide. Lane 1: *C. arietinum* ILC 482; lanes 2-3: *C. arietinum* ILC 1272; lanes 4-5: *C. arietinum* ILC 3279; lane 6: *C. arietinum* ILC 1929; lane 7: *C. reticulatum* ILWC 123; lane 8: *C. reticulatum* ILWC 137; lane 9: *C. echinospermum* ILWC 179. Only low levels of variability are detected. Positions of molecular weight markers are given in kilobases.

1.4. Microsatellite-primed PCR with anchored primers (AMP-PCR)

In a variant of the MP-PCR technique designed by Zietkiewicz et al. (1994), the di- or trinucleotide repeat SSR cores serving as single PCR primers are anchored by short 5'- or 3'-end located nucleotide stretches not included in the repeat sequence. It is therefore designated as **anchored microsatellite-primed (AMP)-PCR**. When anchored at the 5'-terminus, the primers anneal at the distal ends of the SSRs resulting in amplification products that, unlike MP-PCR products, not only contain the regions between the SSRs, but also the SSRs themselves. 5'-anchors have to be designed carefully to prevent slippage towards the 3' end of the microsatellite during PCR (Fisher et al. 1996). If appropriate 5'-anchoring is achieved, any polymorphism of the amplification products is not only originating from the variability of the ISSR sequences, but also caused by the inherent polymorphism of the SSRs themselves. Radiolabeled AMP-PCR amplification products yielded complex banding patterns when separated in polyacrylamide gels and detected by autoradiography (Zietkiewicz et al. 1994; see Fig. 1). These patterns display inter- as well as intraspecific variability (Zietkiewicz et al. 1994; Wolff et al. 1995; Hüttel 1996).

Molecular markers derived from anchored and unanchored MP-PCR have since been used for a variety of purposes, such as genetic mapping (e.g., in Einkorn wheat: Kojima et al. 1998, and chickpea: Ratnaparkhe et al. 1998), cultivar identification (e.g. in potato: Prevost and Wilkinson 1999), the differentiation of species and subspecies (e.g. in *Plantago major*: Wolff and Morgan-Richards 1998), and the assessment of hybridisation in natural populations (e.g. in *Penstemon*: Wolfe et al. 1998). ISSRs were often more polymorphic than RFLPs and RAPDs, for example in finger millet (*Eleusine coracana*) (Salimath et al. 1995). Huang and Sun (2000) used mostly 3'-anchored dinucleotide motifs to determine the genetic relationships of sweet potatoes and related species. A total of 15 primers generated 52 bands per accession on average, of which up to 78 % were polymorphic. In chickpea, pattern complexity and informativeness of AMP-PCR bands were considerably higher than MP-PCR patterns (Fig. 4). While MP-PCR usually resulted in species-specific bands, AMP-PCR also detected intraspecific polymorphisms (Hüttel 1996).

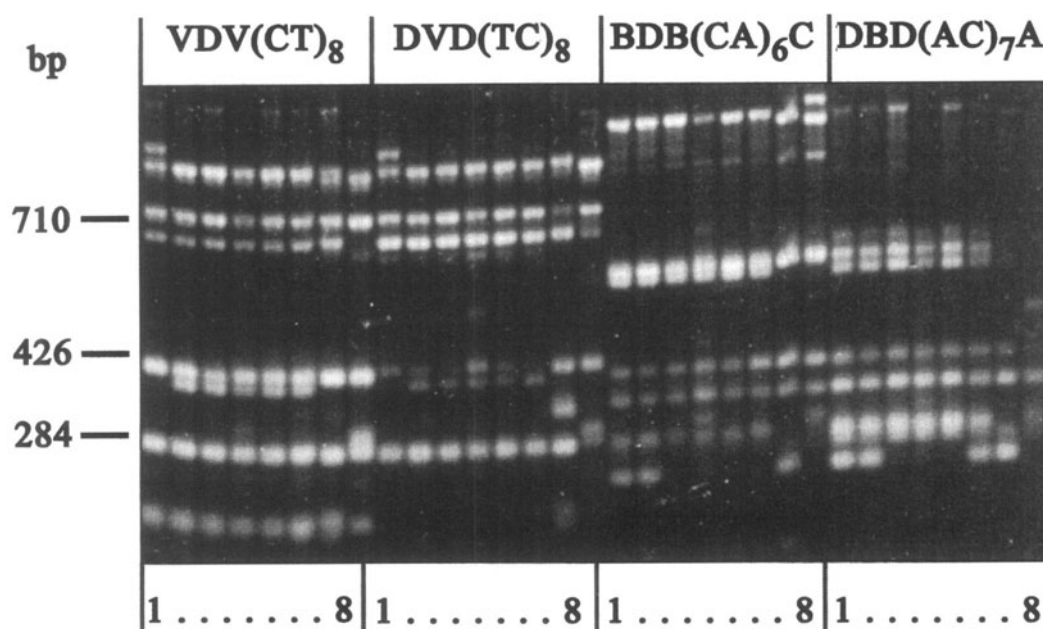


Figure 4: AMP-PCR analysis of intra- and interspecific variation in chickpea. Genomic DNA from individual plants was amplified using the oligonucleotides indicated below the figure as single primers. All primers consist of dinucleotide repeats carrying a degenerate "anchor" at their 5'-ends. PCR products were separated on 1.5% agarose gels and stained with ethidium bromide. Lane 1-6: *C. arietinum* (lane 1: ILC 191; lane 2: ILC 200; lane 3: ILC 482; lane 4: ILC 1272; lane 5: ILC 1929; lane 6: ILC 3279); lane 7: *C. reticulatum* ILWC 36; lane 8: *C. echinospermum* ILWC 35. The level of intra- and interspecific variability is higher as compared to PCR with unanchored primers (see Fig. 3), but lower as compared to RFLP fingerprinting (Fig. 2). Positions of molecular weight markers (in bp) are indicated by arrows. B = C,G or T; D = A,G or T; V=A, C or G.

In summary, AMP-PCR is superior to the unanchored variants of MP-PCR for mainly three reasons. First, the anchoring of primers allows their annealing exclusively at the ends of an SSR, thus obviating internal priming and smear formation. Second, the anchor allows only a subset of the targeted inter-repeat regions to be amplified, thereby reducing the high number of PCR products expected from priming of inter-repeat regions bracketed by dinucleotide repeats to about 10-50 clearly resolved bands. Pattern complexity can then be tailored with different primer lengths and sequences. Third, efficient 5'-anchors ensure that the targeted microsatellite is part of the product which considerably enhances the chance of detecting polymorphisms.

1.5. Combining RAPD and AMP-PCR primers: The RAMP technique

In the "random amplified microsatellite polymorphism" (RAMP) technique developed by Wu et al. (1994), 5'-anchored SSR primers and arbitrary 10mer (RAPD) primers were combined to detect polymorphisms from different *Arabidopsis thaliana* strains and ecotypes. Target loci were amplified by a specifically designed asymmetric PCR protocol that switched between high and low annealing temperatures, favouring amplification from the SSR primer. Since only the SSR primer was radiolabelled with ^{33}P -dATP, autoradiograms of PCR products separated in polyacrylamide gels showed SSR-associated bands only. Ten to thirty polymorphisms per primer (2-7 alleles in 11 different *Arabidopsis* ecotypes) were observed. In a test cross the majority of these fragments segregated as codominant markers (Wu et al. 1994).

Becker and Heun (1995b) located 40 new markers on a barley RFLP map using 5'-anchored, radiolabelled $[\text{GA}]_n$ primers in combination with 10-mer, 16-mer and 20-mer RAPD primers. The use of longer-than-usual RAPD primers allowed comparable annealing temperatures for both types of primer. To obtain additional polymorphisms, aliquots of the amplification products were digested with *Mse*I, resulting in so-called dRAMPs with 0-11 mainly dominant (presence/absence) polymorphisms per primer combination. Ten primer combinations produced 43 RAMPs and 17 dRAMPs. As demonstrated by mapping, the digestion of RAMP products was not efficient, since only 7 dRAMP loci were unique, i.e. the others were actually derived from RAMPs. RAMPs were also used to study the genetic relationships between barley cultivars by Sanchez de la Hoz et al. (1996). There, PCR products were stained with silver, thereby visualizing RAPD-products as well as products derived from both types of primer. Cloning and sequencing of 26 randomly selected RAMP bands from barley showed

various numbers of the expected repeat at the ends of the fragment, as well as the presence of cryptic repeats in the interior of some fragments (Dávila et al. 1999).

In our own RAMP experiments on chickpea, a series of RAPD primers were tested but only few intraspecific polymorphisms were observed. In its present form, RAMP cannot be recommended for genomes with a low level of polymorphism.

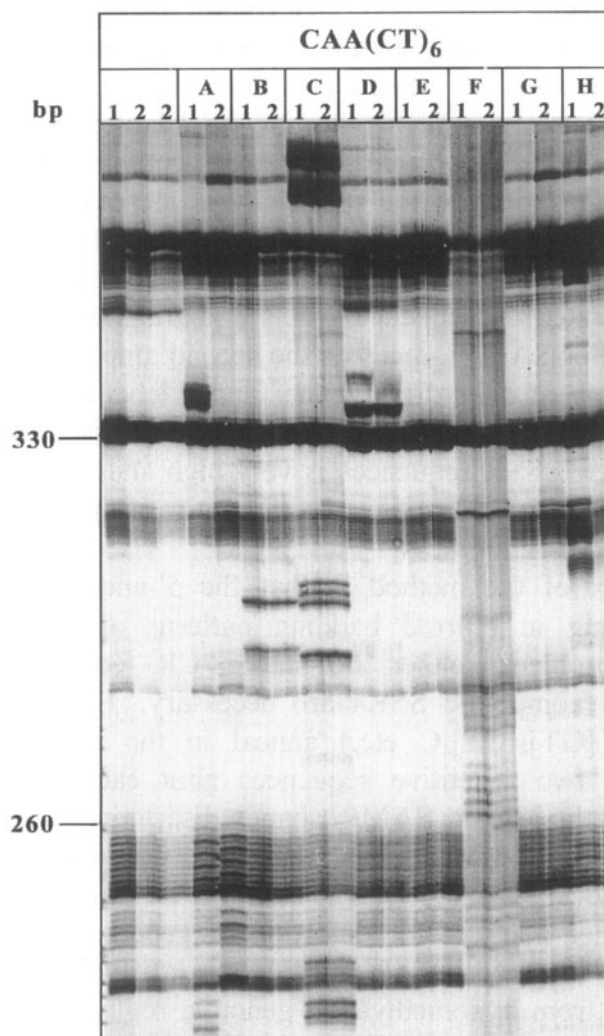


Figure 5: RAMP analysis of chickpea accessions. Genomic DNA samples from individual plants of chickpea accessions C-104 (1) and WR-315 (2) were amplified using the anchored SSR primer CAA[CT]₆ either singly or in combination with various RAPD primers (A = OPAQ-02; B = OPAQ-07; C = OPAQ-09; D = OPAQ-13; E = OPAQ-14; F = OPAQ-17; G = OPAQ-18; H = OPM-01; all from Operon Technologies, Alameda, USA). The anchored SSR primer was ³³P-labeled by T4 polynucleotide kinase. PCR products were separated on a denaturing 6% polyacrylamide gel and autoradiographed. Bands that are visible throughout *all* lanes are derived from the SSR primer only, bands specific for either A, B, C, D, E, F, G or H are derived from mixed priming. Only few polymorphisms between the two investigated chickpea accessions are detected. Positions of molecular weight markers are given in bp.

1.6. Selective amplification of microsatellite polymorphic loci (SAMPL)

The SAMPL technique combines the controllable multiplexing rate of the amplified fragment length polymorphism (AFLP, Vos et al. 1995) technique with the presumably higher levels of SSR polymorphism by utilising AFLP-type primers together with AMP-PCR primers (Vogel and Scolnik 1997). In short, DNA is digested with one rare and one frequently cutting restriction enzyme and suitable adapters are ligated to the resulting restriction fragments. For the detection of standard AFLPs, the DNA is first amplified from primers complementary to the adapter and to one specific base extending from the 3'-end into the ligated genomic DNA fragment. In a second round of amplification the specificity is increased by using up to 3 selective bases at the primer's 3'-end to reduce the total number of fragments amplified. Only one of the primers is labelled with ^{33}P or ^{32}P . Amplification products are separated on sequencing gels and fragments are visualized by autoradiography. For SAMPL analysis the second amplification is performed using a combination of one AFLP primer and a labelled 5'-anchored AMP-PCR primer. Since the number of possible combinations of restriction enzymes and targeted SSR sequences is very high and each combination can produce from 10 to 50 bands, the amount of potentially detectable polymorphisms is nearly limitless.

A drawback of the method is that the 5'-anchor is frequently not functional, resulting in blurred banding patterns and stutter bands. To overcome this problem, either hot-start PCR conditions or primers complementary to compound SSRs are necessary. These primers (as e.g. A[CA]₇[TA]₂T, T[GT]₄[CT]₄C etc.) anneal in the interior of compound SSRs, where the two repetitive sequences abut each other, providing a perfect anchor for the primer. SAMPL bands discriminated between soybean cultivars and species, and Mendelian inheritance was demonstrated for most of them (Vogel and Scolnik 1997). The SAMPL technology has been used in comparison to other markers for detection of genetic diversity in pea and performed similar to standard AFLPs (Lu et al. 1996). More recently, 20 SAMPLs derived from hypomethylated genomic regions have been mapped onto the genome of Norway spruce (*Picea abies*), together with 366 AFLPs and 61 SSRs (Paglia et al. 1998). All SAMPL bands behaved as dominant markers.

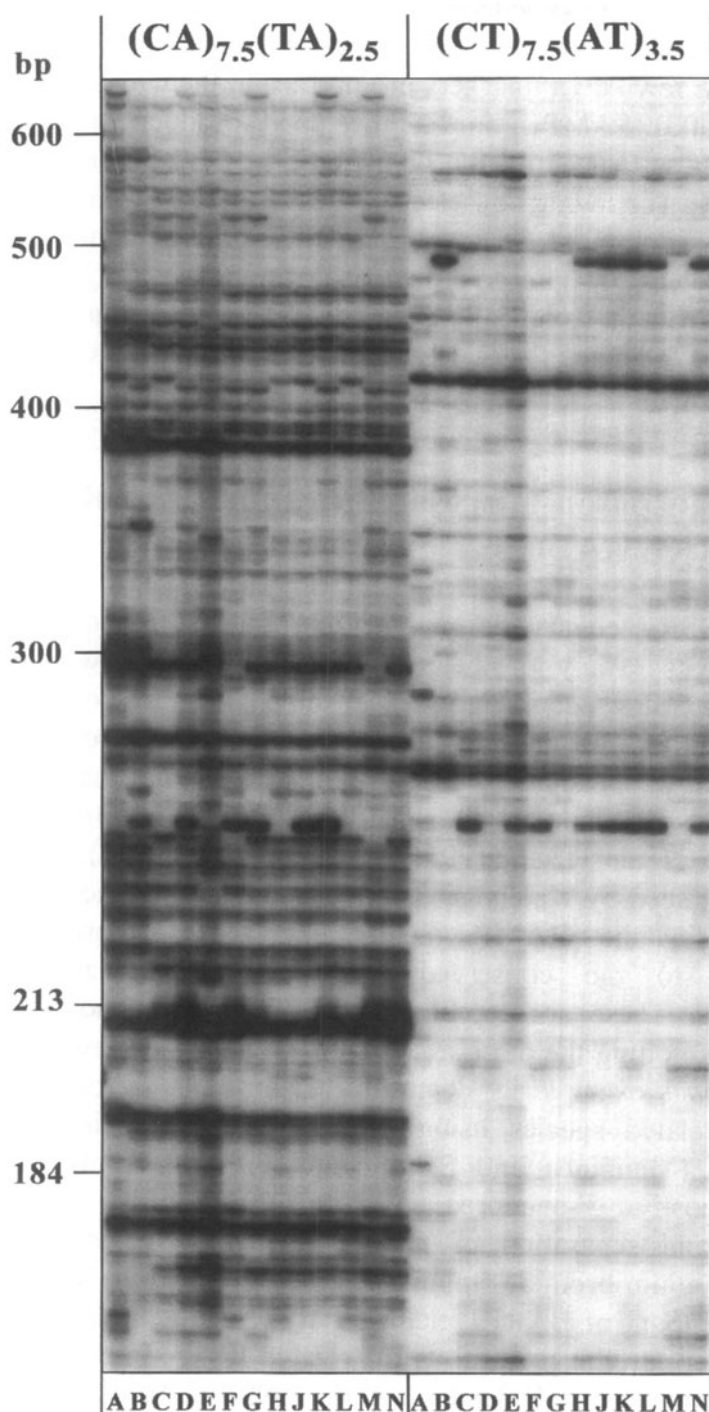


Figure 6: SAMPL analysis of recombinant inbred lines (RILs) from a *C. arietinum* x *C. reticulatum* interspecific cross. For SAMPL analysis, plant DNA was first digested and then an *Eco* RI restriction site-specific adaptor was ligated to the fragments. The ^{32}P -endlabelled compound SSR-primer indicated above the autoradiograph was used in combination with an primer complementary to the *Eco* RI adaptor. PCR products were separated on 4% standard sequencing gels. Positions of co-electrophoresed ^{32}P -endlabelled DNA markers are given in bp. A = *C. reticulatum* (PI489777), B = *C. arietinum* (ICC4958), C to N = eleven randomly selected recombinant inbred lines.

In chickpea, the low frequency of RFLPs on which AFLPs are based precluded the large-scale exploitation of this technique for genome mapping. Therefore, only a few AFLP markers are present in the current chickpea map (Winter et al. 2000). SAMPL, on the other hand, produces large numbers of polymorphic bands. Especially the combination of Eco RI-adaptor primers of low selectivity (i.e. with no selective bases at the 3'-end) and primers directed against compound SSRs results in complex, highly polymorphic patterns, that are easy to interpret because of their clarity (Fig. 6). Despite the fact that SAMPL is technically demanding it is the dominant, SSR-based marker technique of choice for chickpea and will speed up the mapping of its genome tremendously.

1.7. Retrotransposon-microsatellite amplified polymorphisms (REMAP)

Besides SSRs, eukaryotic genomes harbour a large number of other repetitive elements. Many of these are retrotransposons which are often accompanied by SSRs, occurring either at the 5' or 3'-end, or in the interior of the transposon (e.g., Ramsay et al. 1999). The frequent colocalization of both elements led to speculations that both coevolved, and that SSRs provide integration points for homology-driven insertion of retrotransposons into genomic DNA (Nadir et al. 1996; Ramsay et al. 1999). The REMAP technique, introduced by Kalendar et al. (1999), replaces the random 10mer primers used in the RAMP method by outward oriented primers complementary to the conserved long terminal repeats (LTRs) of retrotransposons. These are combined with 3'- or 5'-anchored AMP-PCR primers, thus exploiting the close association of SSRs with retrotransposons. The technique was exemplarily demonstrated on DNA of barley (*Hordeum vulgare*) and related species using primers directed against the BARE1 retrotransposon (Manninen and Schulmann 1993) in combination with a variety of mostly 3'-anchored di- or trinucleotide SSR primers. Amplification and separation of products in 2% NuSieve agarose gels followed by ethidium bromide staining resulted in 15 to 30 bands, reflecting the estimated amount of BARE1 elements in the different *Hordeum* species. REMAP was similarly successful as the related IRAP (inter-retrotransposon amplified polymorphism, Kalendar et al. 1999) technique in that bands generated by both methods were 100% polymorphic between the species, and could also differentiate below the species level. A similar strategy, coined *Copia*-SSR was simultaneously developed by Provan et al. (1999b) and also tested on barley. Multiple polymorphic products were obtained, and seven markers were mapped on four different chromosome arms.

Probe	Clones	Repeat Motif	Number of Sequences	Number of Repeat Units	Number of Perfect Repeats
[GA] ₈	26	GA	13	9-32	10
[GT] ₈	18	GT	5	4-42	3
[TAA] ₅	39	TAA	13	5-54	10
		AT	5	9-20	3
A/T-pool	27	GAA	1	9	1
		CAA	3	4-7	3
		CAT	1	5	1
G/C-pool	11	CAG	2	4-5	2

Table 1: Pilot study for the generation of STMS markers in chickpea (Hüttel et al. 1999)

2. SPECIFIC AMPLIFICATION OF MICROSATELLITE LOCI FROM FLANKING PRIMERS: STMS MARKERS

Currently, the most popular method to exploit SSR variability for the generation of genetic markers uses SSR-flanking sequences as primers to amplify the enclosed SSR. The resulting locus-specific amplification products often exhibit considerable length differences among different individuals or populations of the same species, mostly due to the variable number of tandem repeats within the SSR. These "sequence-tagged microsatellite site (STMS, Beckmann and Soller 1990) markers are the markers of choice for nearly every organism. They represent single-locus, co-dominant, easy-to-use and reliable markers with high polymorphic information content possessing the potential for automated, non-radioactive detection

2.1. History of STMS technology

Locus-specific SSR PCR was introduced in 1989 by four groups independently (Smeets et al. 1989, Tautz 1989, Litt and Luty 1989, Weber and May 1989). The term "microsatellite" was created by Litt and Luty (1989), and the whole technology of marker generation was coined STMS analysis by Beckmann and Soller (1990). Since then, STMS markers were successfully used to generate high-density marker maps of the human (Weissenbach et al. 1992, Dib et al. 1996) and many animal genomes (e.g. Jacob et al. 1995; Dietrich et al. 1996). In 1992 these elite markers were first introduced to plant genome analysis (Akkaya et al. 1992).

Considerable numbers of STMS markers are now available for many crop plants and model species such as *Arabidopsis* (Bell and Ecker 1994;), banana (Kaemmer et al. 1997), barley (Saghai-Maroo et al. 1994, Becker

and Heun 1995a, Liu et al. 1996), chickpea (Hüttel et al. 1999, Winter et al. 1999), *Citrus* (Kijas et al. 1995; 1997), grapevine (Thomas et al. 1994, Scott et al. 2000), kiwifruit (Weising et al. 1996, Huang et al. 1998), lettuce (Van de Wiel et al. 1999), maize (Taramino and Tingey 1996, Chin et al. 1996), *Pelargonium* (Becher et al. 2000), Norway spruce (Pfeiffer et al. 1997), potato (Milbourne et al. 1998), rice (Wu and Tanksley 1993, Akagi et al. 1997, McCouch et al. 1997, Cho et al. 2000), soybean (Morgante et al. 1994, Akkaya et al. 1995, Diwan and Cregan 1997), tomato (Smulders et al. 1997), tropical trees (Chase et al. 1996; White and Powell 1997a,b), wheat (Röder et al. 1995, 1998; Bryan et al. 1997, Stephenson et al. 1998), or yams (Terauchi and Konuma 1994), to name only a few. In spite of the many advantages of STMS, some limitations also have to be considered. The main obstacles are the high costs for cloning, sequencing and primer synthesis. Further, standard protocols still use radioisotopes and sequencing gels to detect the amplified SSRs. Finally, the efficiency of primer generation suffers from a number of problems including redundancy of clones and the occurrence of artificial chimeras.

2.2. Generation of STMS markers

A prerequisite for the generation of STMS markers is knowledge of SSR-flanking sequences which is either available from databanks or from sequencing of cloned, SSR-containing fragments. These can be obtained either by directly screening small-insert genomic libraries with suitable probes, or by enrichment of the libraries for SSRs prior to hybridization and sequencing. In the following we will discuss some of the available cloning and enrichment strategies as well as several aspects that have to be considered before starting the work.

The generation of STMS markers generally requires the following steps: (1) Isolation of genomic DNA; (2) digestion with one or more 4 base-specific restriction enzymes (or sonication); (3) size-selection of restriction fragments by agarose gel electrophoresis, excision and purification of the desired size fraction; or alternatively, enrichment of the fraction of SSR-containing fragments by preselection and PCR; (4) ligation of the DNA into a suitable vector and transformation into *E. coli*; (5) screening for the presence of SSRs by colony- or plaque-hybridization with a labeled probe; (6) isolation of positive clones and sequencing of the inserts; (7) design of suitable primers flanking the repeat.

Though the production of STMS markers is expensive in terms of time and money, their generation is worth the effort. However, some points should be considered before embarking on such a project.

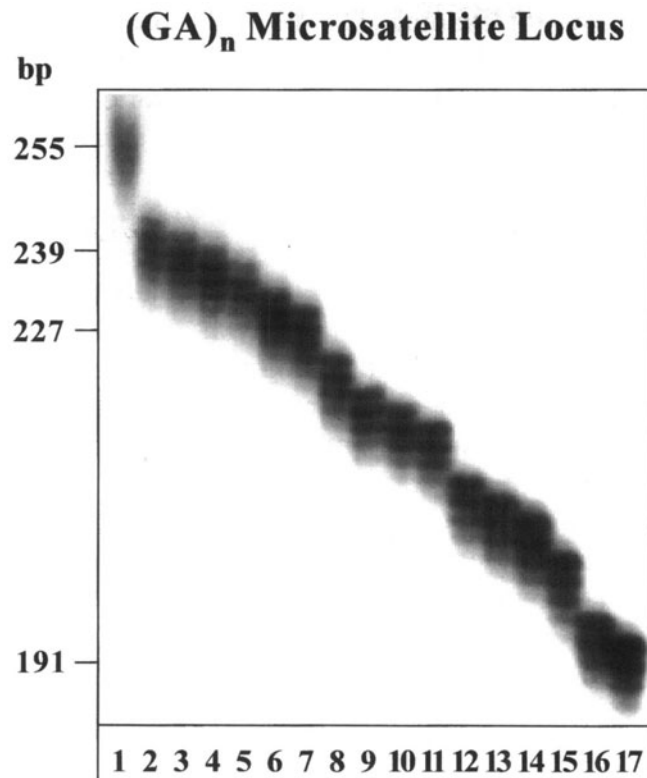


Figure 7: Detection of large numbers of alleles at a $[\text{GA}]_n$ SRR locus by STMS analysis. Only a subset of alleles present in 63 selected *C. arietinum* accessions as revealed by locus-specific amplification from a STMS primer pair is shown. ^{33}P -labelled PCR products were run on standard sequencing gel and visualized by autoradiography. To facilitate the determination size of product sizes, products were arranged by size and run on a second gel. Lanes 1-17: ILC 2564 (1), ILC 6152 (2), ILC 2 (3), ILC 374 (4), ILC 6092 (5), ILC 6102 (6), ILC 236 (7), ILC 5899 (8), ILC 2548 (9), ILC 4328 (10), ILC 5810 (11), ILC 37 (12), ILC 5871 (13), ILC 188 (14), ILC 1994 (15), ILC 1272 (16), WR-315 (17). Fragment sizes were calculated by comparison with a co-electrophoresed M13 ssDNA sequencing ladder.

2.2.1. Initial considerations

If sequence information of SSRs is already deposited in data bases for the crop in question or for related crops, it should be used. Especially for model organisms like *Arabidopsis*, and agronomically important plants like rice or potato, sufficient genomic data are available that can be screened for SSR motifs (Smulders et al. 1997; Milbourne et al. 1998, Cho et al. 2000). Moreover, for several plant species including *Arabidopsis* (Delseny et al. 1997), maize (Wang and Bowen 1998), rice (Yamamoto and Sasaki 1997), or grape (Scott et al. 2000), expressed sequence tag (EST) databases have been developed from which SSR-containing sequences in and around expressed genes can be extracted and used for the design of primers. In the

grape EST database, for example, 2.5 % of the presently 13,000 sequences contain SSRs of the di- and trinucleotide type (Scott et al. 2000).

If, as is the case for chickpea, such information is not available, the SSR motifs and the method for selecting SSR-containing fragments should be carefully chosen. Informativeness is one important criterion. Generally, SSRs obtained from genomic libraries appear to be more polymorphic than those derived from EST databases (Cho et al. 2000, Scott et al. 2000). The latter are often shorter (less than 10 repeat units) than SSRs cloned at random from genomic libraries (Temnykh et al. 2000). Also, the discrimination power of SSRs from 3'-untranslated regions (UTRs) of ESTs is higher than those in 5'-UTRs, whereas SSRs from coding regions often differentiate only between species and genera (which makes them useful at higher taxonomic levels; Scott et al. 2000).

In accordance with the slipped-strand mispairing mutation model of SSRs (Levinson and Gutman 1987), a strong positive correlation between the number of perfect, uninterrupted SSR repeats on the one hand, and the number of alleles on the other is often observed (Weber 1990, Saghai-Maroo et al. 1994, McMurray 1995, Wierdl et al. 1997, Brinkmann et al. 1998). At a first glance, for the generation of informative markers it seems therefore preferable to select SSR types consisting of long, perfect arrays. For example, in the tea tree, perfect trinucleotide repeats exhibited nearly twice as many alleles as imperfect ones (Rosetto et al. 1999). However, long SSRs may also be disadvantageous in some situations. In tomato, Areshchenkova and Ganai (1999) have shown that long (>20 repetitions) arrays of GA-, AT-, CA- and GATA repeats are predominantly clustered at centromeric regions, and hence, are quite useless for genetic mapping.

The percentage of perfect vs. imperfect and/or compound motifs varies considerably among different studies, sometimes even in the same species. In their study on lettuce, Van de Wiel et al. (1999) noted that a smaller proportion of imperfect microsatellites were obtained after enriching the genomic libraries with a single microsatellite motif rather than with mixtures of several motifs. A similar observation was made in *Pelargonium*, where a lot more perfect repeats were obtained by enriching the library by hybridization with a pure GA repeat, as compared to a mixture of CA-, GAA- and CAA- repeats (Becher et al. 2000). It is possible that the use of oligonucleotide mixtures for enrichment could result in the selective accumulation of compound repeats.

Mutation rate and informativeness of a locus may also depend on the number of nucleotides in the basic repeat unit. At least in man and *Drosophila*, mutation rates of dinucleotide repeats are higher than those of most tri- or tetranucleotide arrays (Chakraborty et al. 1997, Schug et al. 1998). Whereas EST-derived GC-rich trinucleotide repeats from rice (i.e. [GCC] and [GAC]) were generally less polymorphic than dinucleotide repeats, AT-rich trinucleotide repeats from genomic libraries were as variable

as the most polymorphic GA repeats from the same source (Cho et al. 2000). This demonstrates that trinucleotide and other SSRs form a heterogeneous group of loci, each of which may have its own motif-, locus- or even allele-specific mutation potential. Another criterion is the genomic abundance of a particular SSR type. Though A-, TA- and TAA repeats outnumber other SSRs in many plants, this may not be true for every crop. For example, in rice and maize [CGG]_n arrays are most prevalent (Zhao and Kochert 1993, Panaud et al. 1995, Chin et al. 1996), followed by [ACG]_n and [AG]_n repeats in rice (Temnykh et al. 2000).

Yet practical problems also have to be considered. For example, tri- and tetranucleotide repeats are probably better suited as markers than mono- or dinucleotide motifs, because of the occurrence of "stutter" bands often visualized on sequencing gels. Instead of one particular band, the enzymatic amplification of mono- and dinucleotide repeats commonly yields a cluster of "shadow bands", which result from slippage events during replication catalyzed by *Taq* polymerase (Hauge and Litt 1993). Slippage of amplified tri- and tetrameric repeat units is usually less severe than that of mono- or dinucleotide repeats (Edwards et al. 1991, Hearne et al. 1991, Kijas et al. 1995). Moreover, allele size differences of one or two tri- or tetranucleotide repeat units can be clearly resolved on native polyacrylamide gels or even agarose gels, which is not the case for mono- or dinucleotide repeats. Stutter bands may make interpretation of correct allele sizes difficult or even impossible, especially if two alleles differ by one or two base pairs only. Ordering of alleles according to size may help to correctly assign allelic states (Saghai-Maroo et al. 1994).

For all of these reasons, a pilot study in which a small set of different SSR types are isolated and sequenced, should identify the optimal motifs for a particular crop. For example, in our chickpea pilot study 13.000 clones were obtained from two small insert libraries generated with a mixture of 7 different blunt-end producing restriction enzymes that contained size-selected fragments of 250 to 400, and 400 to 600 bp, respectively (Hüttel et al. 1999). Establishing libraries with small, size-selected inserts is advantageous for SSR cloning, because very long microsatellites are unstable in *E. coli*, and cloned inserts can be sequenced without subcloning. Also, most chimeric clones can be identified by their size. The chickpea libraries were screened with [TAA]₅, [GA]₈, [GT]₈, pools of A/T-rich trinucleotide repeats ([CAA]₅, [CAT]₅ and [GAA]₅), and pools of G/C-rich trinucleotide repeats ([TCC]₅, [CAC]₅, [CAG]₅ and [CGA]₅), respectively. These identified a total of 121 positive colonies, of which 51 were sequenced. As is shown in Tab. 1, [TAA] repeats were the most abundant type of repeat in chickpea. Since [TAA] arrays were often quite long, and mostly perfect (Hüttel et al. 1999), this type of repeat offered itself for the generation of large numbers of STMS markers in chickpea.

An important point to be clarified, before choosing a particular SSR motif for large-scale marker generation is the genomic distribution of the repeats. This could be done either by in-situ hybridization or by genetically mapping some of the loci in early phases of the project. For instance, in *Arabidopsis* SSRs of the G/A type (Brandes et al. 1997) and in sugar beet of the G/T-type are part of abundant repetitive elements crowded near centromeres (Schmidt and Heslop-Harrison 1996). The same is true for long [GATA]_n, [AT]_n, [GA]_n and [GT]_n arrays in tomato (Areshchenkova and Ganai 1999; see above). Also in chickpea metaphase spreads, many [CAA] repeats proved to be clustered (Gortner et al. 1998). These microsatellite types obviously are only second choice for genome mapping.

Technique	Costs	Efforts	Polymorphism		Suitability
			intraspecific	interspecific	
RFLP fingerprinting	low-medium	high	high	high	low
RAMPO	low-medium	high	medium	medium	low
MP-PCR	low	low	low	medium	low
AMP-PCR	low	low-medium	low-medium	medium-high	medium-high
RAMP	low	low-medium	low	low-medium	low
REMAP	low	low	low	low	low
SAMPL	low	low-medium	medium	high	high
STMS	high	low-medium	high	high	high

Table 2: Suitability of microsatellite-based marker techniques for chickpea

2.2.2. Large-scale generation and application of STMS markers: Allele sizing and linkage mapping

As an immediate consequence of the pilot study of Hüttel et al. (1999), a size-selected genomic library comprising 280,000 colonies and representing about 18 % of the chickpea genome, was screened for clones containing the most abundant [GA]_n, [GAA]_n and [TAA]_n -motifs (Winter et al. 1999). A total of 389 positive colonies were sequenced. The majority (~ 75 %) contained perfect repeats. Interrupted, compound and interrupted-compound repeats were only present from 6 % to 9 %. As expected from the pilot study, fragments isolated by the [TAA] probe contained the longest repeats with unit numbers from 9 to 131 (see also Fig. 8).

For 218 loci, primers could be designed, and were used for the detection of length polymorphisms among eight *Cicer* genotypes, including 6 relevant chickpea breeding cultivars and *C. reticulatum* and *C. echinospermum*, chickpea's wild, intercrossable relatives. A total of 174 primer pairs gave interpretable banding patterns, 137 (79 %) of which revealed at least 2 alleles on native polyacrylamide gels.

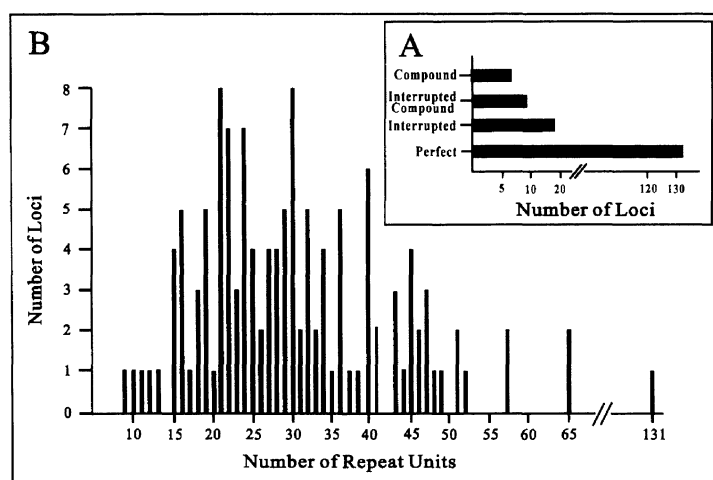


Figure 8: Number of compound, interrupted compound, interrupted and perfect [TAA]_n-SSR arrays (B), and number of TAA repeat units in perfect SSRs from chickpea (A).

Of these, 120 were genetically mapped on 90 recombinant inbred lines from an inter-species cross between *C. reticulatum* and the *C. arietinum* cultivar ICC 4958. Markers could be arranged in 11 linkage groups (LOD-score: 4) covering 613 cM. Most markers were distributed randomly, but some local clustering was observed. For 46 markers (39 %), segregation deviated significantly ($P > 0.05$) from the expected 1:1 ratio. The majority of these aberrantly segregating loci (73 %) were located in three distinct regions of the genome. STMSs have been used for genome mapping in many other crops (review in Gupta and Varshney 2000). STMS marker maps of similar size or even more extended than in chickpea are available for e.g. barley (Liu et al. 1996), potato (Milbourne et al. 1998), rice (McCouch et al. 1997, Chen et al. 1997, Temnykh et al. 2000, Lorieux et al. 2000), soybean (Cregan et al. 1999a) and bread wheat (Stephenson et al. 1998, Röder et al. 1998, Korzun et al. 1999).

2.3. Exploiting the codominant nature of STMS markers: Transfer between populations

Their extensive degree of polymorphism set aside, the major advantage of single-copy STMSs is their codominant nature, that allows the detection of heterozygotes in F₂-populations and their transfer between populations. For breeders, the ease of handling and robustness of STMS make them especially useful as anchor markers. STMS can be used to confirm a map position obtained in one segregating population, in another one, which is a prerequisite for marker-assisted selection (MAS). For many important crops like bread and durum wheat, rice, and soybean (see review in Gupta and Varshney 2000), and also in chickpea (Rathnaparke et al. 1998,

Santra et al. 1998, Winter et al. 2000), SSR markers are now available that tag resistance loci and other genes of agronomical importance.

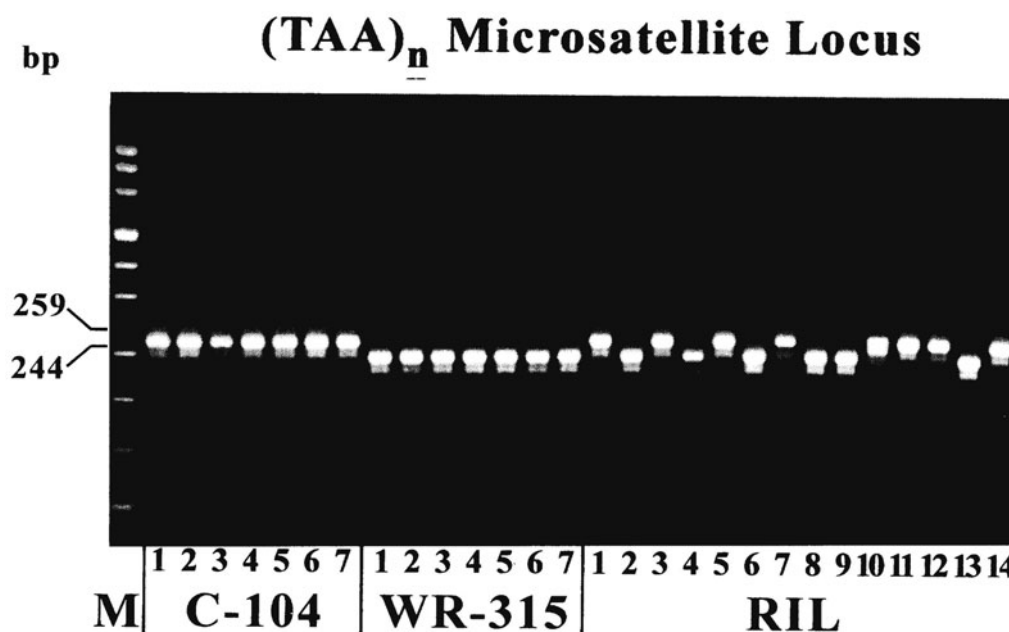


Figure 9: Mendelian inheritance of a microsatellite locus in recombinant inbred lines (RILs) of chickpea. Seven plants randomly selected from either parental accession and fourteen RILs were analyzed with a primer pair flanking a (TAA)_n locus. PCR products were electrophoresed in 2.5% Metaphor gels and stained with ethidium bromide. M: size marker. Sizes of parental alleles evaluated on sequencing gels are indicated on the left.

However, for many genomes the number of available STMS markers is still too small to allow the tagging of specific genes, or - more desirable - chromosome landing, i.e. the direct isolation of genes from YAC- or BAC libraries without any "walking" steps (Tanksley et al. 1995). In these cases it is advised to combine STMS with dominant markers such as RAPD, DNA amplification fingerprint (DAF, Caetano-Anolles et al. 1991) or amplified fragment length polymorphism (AFLP, Vos et al. 1995) markers or their variants. For example, in an effort to localize the major gene(s) conferring resistance to *Ascochyta rabiei*, an important fungal chickpea pathogen, on the genomic map, we combined DAF with STMS and bulked segregant analysis (BSA, Michelmore et al. 1991) to first tag the locus with DAF markers in a population segregating for this locus. The two DAF markers OPS06-1 and OPS03-1 were linked to the resistance locus at LOD-scores > 5. OPS03-1 could be transferred to the population on which the current genetic map of chickpea is based (Winter et al. 2000), where it mapped to linkage group 4. STMS markers localised on that linkage group were in turn transferred to the population segregating for *Ascochyta* resistance. Three of these markers were closely linked to the major resistance locus, and twelve out of 14 STMS markers from linkage group 4 could be used in both

populations. The order of STMS markers was essentially similar in both populations with some differences in map distances (Rakshit et al. in preparation). In future, the availability of STMS markers for *Ascochyta* resistance will certainly help to decipher the complex reactions of the chickpea host against different pathotypes of the fungus.

2.4. Targeted isolation of STMS markers from large-insert clones

Experiences with the high-density STMS marker map of soybean containing more than 412 markers revealed that, despite the large number of markers, still 36 intervals of 20 cM each are not covered by STMS (Cregan et al. 1999a). In these cases, or when markers are not at all available for MAS, STMS markers for a specific gene can be directly isolated from large-insert yeast (YAC) or bacterial (BAC) artificial chromosome libraries using other, closely linked markers as starting points. YAC and BAC vectors often contain fragments of 100 kb or more in size, which are good substrates to find one or more SSRs since these are so abundant in plant genomes.

Several techniques exist for the screening of large-insert libraries with DNA markers which often depend on pooling strategies (see e.g. Marek and Shoemaker 1997). Once a large-insert clone has been identified, the YAC or BAC is isolated, fragmented by restriction enzymes, and the fragments - small enough to allow their direct sequencing - subcloned into suitable vectors. Small-insert libraries are then screened by hybridization with SSR-specific oligonucleotides, positive clones sequenced and STMS primers designed. Using this strategy, Cregan et al. (1999b) developed three new $[AT]_n$ and $[ATT]_n$ -containing STMS markers each for two regions of the soybean genome harboring resistance genes for the soybean cyst nematode.

To circumvent the tedious cloning and screening steps, Lench et al. (1996) combined 3'-anchored AMP-PCR primers with an unspecific vectorette primer (Riley et al. 1990) to develop STMS markers for the human amelogenin gene, starting from cosmid and YAC clones. These large inserts were digested with different restriction enzymes, ligated to the vectorette and amplified after annealing of the microsatellite and vectorette-specific primers. SSR-harboring PCR products were cloned and sequenced, and the sequence information from one flank of the microsatellite was used to produce a specific primer. This primer and the vectorette primer then served for the amplification of the other flank of the respective microsatellite. After sequencing, the second primer was designed. Lench et al. (1996) were able to generate six polymorphic STMS markers for the amelogenin gene that way.

If a targeted STMS is going to be prepared from a YAC clone, the above method requires the isolation of the large YAC fragment from which the microsatellite shall be cloned, because yeast DNA also contains SSRs and thus also serves as template for the AMP-PCR primer. This problem can be overcome by replacing the vectorette primer by a primer selective for species-specific repetitive elements (interspersed repetitive sequences, IRS, Phan et al. 2000). The specificity of the IRS primers prevent to amplify yeast DNA, and hence, the PCR products are exclusively derived from the cloned DNA. As above, the PCR products derived from SSR-IRS PCR are cloned, sequenced and a flanking primer designed, directing outwards of the IRS sequence. The YAC-containing yeast DNA is then digested separately with different restriction enzymes, a vectorette ligated to the fragments, and PCR performed with both the newly designed primer and the vectorette primer. The reaction product is again sequenced and another, specific STMS primer designed for the other flank of the product. This technique produced 8 STMS from a 600 kb YAC (Phan et al. 2000). Though it was developed for mouse YACs originally, it can probably be applied to plant DNA, since species-specific interspersed elements such as the BARE1 retrotransposon (see above) are abundant in plants as well.

2.5. Transferability of STMS markers within genera: Conservation and variability of STMS from chickpea within the genus *Cicer*

Currently, one major limitation of STMS technology is the species specificity of markers. Unless sufficient database information is available, microsatellites have to be cloned and sequenced for every species under study. These requirements render the establishment of STMS markers for a new species quite cumbersome. The situation would be greatly improved, if primer binding sites and microsatellite loci were conserved among different taxa. The need for marker-assisted exploitation of the primary and secondary gene pool of crops for quality improvement has triggered research that aims at testing such conservation of priming sites and the transferability of STMS markers from one species to another. In general, the extent of STMS marker transferability among species will depend on the evolution rate of the microsatellite-flanking sequences as well as of the microsatellites themselves. Studies undertaken in this direction so far have shown that these evolution rates are highly variable, and depend on the locus as well as on the taxa investigated.

In animals it was frequently observed that SSR-flanking sequences are conserved in closely related species. For example, human STMS primers amplified the corresponding loci from chimpanzees (Deka et al. 1994,

Rubinsztein et al. 1995, Garza et al. 1995), bovine primers detected alleles from goats and sheep (Moore et al. 1991, Forbes et al. 1995, Pepin et al. 1995), and mouse primers those from rats (Kondo et al. 1997). Transferability of markers across families or even orders was observed in e.g. *Canidae* (Roy et al. 1994) and birds (Primmer et al. 1996). The most extreme cases of primer target conservation have been encountered in marine species, where microsatellite loci were shown to be conserved in whales, turtles and fish over several hundred million years of evolution (Schlötterer et al. 1991, FitzSimmons et al. 1995; Rico et al. 1996).

The situation is less promising in the plant kingdom, where STMS marker transferability seems to be mainly limited to congeneric species, as has been shown in e.g. *Citrus* (Kijas et al. 1995), *Actinidia* (Weising et al. 1996; Huang et al. 1998), *Pinus* (Karhu et al. 2000, Echt and May-Marquardt 1997), *Medicago* (Diwan et al. 1997), *Glycine* (Peakall et al. 1998), *Prunus* (Cipriani et al. 1999) or grapes (Scott et al. 2000). Occasionally, markers also amplify fragments from other genera belonging to the same family (e.g. *Citrus* and *Poncirus*: Kijas et al. 1995; Mimosaceae: Dayanandan et al. 1997, Meliaceae: White and Powell 1997b; potato and tomato: Smulders et al. 1997; Brassicaceae: Van Treuren et al. 1997, Westman and Kresovich 1998). From a practical point of view, STMS marker transfer can only be called "successful" or "informative" if the products are *polymorphic* within source as well as non-source species. In other words: not only the primer binding sites, but also the microsatellite itself should be conserved (and hopefully polymorphic). Whether this is the case, can be analyzed by either of three methods: (1) Southern or dot blot analysis of heterologous amplification products, (2) comparative sequence analysis, and (3) screening for the extent of polymorphism within the non-source species. Studies in this direction have shown that in general, only a small subset of amplification products actually contains a repeat in the heterologous taxon, and the percentage of polymorphic loci rapidly drops with increasing phylogenetic distance (Whitton et al. 1997, Westman and Kresovich 1998, Peakall et al. 1998). Mechanistically, the decrease of polymorphism observed in non-source species is probably caused by the interruption of long, contiguous microsatellite arrays by base substitutions, and/or by shortening of perfect arrays through slippage events

Systematic surveys of STMS marker transferability in large plant families have been reported from e.g. sunflower (Asteraceae; Whitton et al. 1997) and soybean (Fabaceae, Peakall et al. 1998). Both studies demonstrated that the majority of markers work well within the genus of origin (i.e. *Helianthus* and *Glycine*, respectively), but only a very small proportion of markers is transferable to more distant members of the family. Peakall et al. (1998) also observed size homoplasy in that alleles of identical size obtained from different taxa differed in their sequence. Hence, SSRs and their flanking regions appear to be less conserved in plants than in animals,

and the transferability of STMS markers across genus borders is quite limited. The reasons for this are enigmatic. Differences in general mutation rates, repair systems, genome organization, life history and selective constraints may play a role. To explain the unusually high level of sequence conservation in marine animals, it was suggested that aquatic environments may be less mutagenic than terrestrial ones (for a detailed discussion see Schlötterer et al. 1991; Rico et al. 1996). One possibility to increase informative STMS marker transferability among divergent plant taxa may be the exploitation of cDNA and EST libraries for marker generation, because SSRs and their flanking sequences present in exons are more likely to be conserved. Recent studies using EST libraries suggest that this may indeed be the case (Scott et al. 2000).

The general trend described above also applies to our own study on STMS primer transferability within the genus *Cicer*. Efforts to improve the agricultural value of chickpea by marker-assisted breeding resulted in the generation of more than 200 STMS markers, many of which have already been used for the generation of genetic maps of the crop (Winter et al. 1999). For ninety of these markers it was explored whether and to which extent they could also be applied to genome analysis of wild *Cicer* species.

The genus *Cicer* comprises 9 annual and 33 perennial species classified into 4 sections (van der Maesen 1987). Evolutionary and genetic relationships of the eight annual species grouped into section *Monocicer* are well described. *C. chorassanicum*, the ninth known annual species, has been classified into section *Chamaecicer*.

Annual species have been subdivided into 4 groups on the basis of crossability, karyotype, isozyme polymorphisms and seed storage protein characteristics. The first group contains the cultigen, its presumable ancestor *C. reticulatum*, and *C. echinospermum* which produce fertile hybrids in crosses with the cultigen, and the perennial *C. anatolicum*. The second group comprises *C. bijugum*, *C. pinnatifidum*, *C. judaicum* and *C. yamashitae*. The remaining species, *C. cuneatum* and *C. chorassanicum*, make up the third and fourth group. We tested whether conservation of SSR-flanking sequences would reflect the known evolutionary relationships between these species. The DNAs derived from five accessions each of the eight annual *Cicer* species and one accession of *C. anatolicum* were amplified with primers derived from 90 SSR-flanking sequences from chickpea. Further, we exemplarily investigated what might be the reason(s) for the differences in number and size of amplification products derived from the same or different species.

In general, the primer sequences successfully amplified PCR products in the related species, indicating conservation of most SSR-flanking sequences among chickpea's relatives. As shown in Fig. 10, size and number of amplification products varied considerably between and within species. The extent of primer site conservation in the various *Cicer* species was in

accordance with their known phylogenetic relationship(s) to chickpea. It ranged from 92.2 % in *C. reticulatum*, chickpea's closest relative and potential ancestor, to 50 % in *C. cuneatum*. A phylogenetic tree, constructed on the basis of presence or absence of amplification products revealed a close relationship between cultivated chickpea and its crossability group on one side, and the perennial *C. anatolicum* on the other. Both groups were more closer to each other than to other annual species of the genus (Fig. 10).

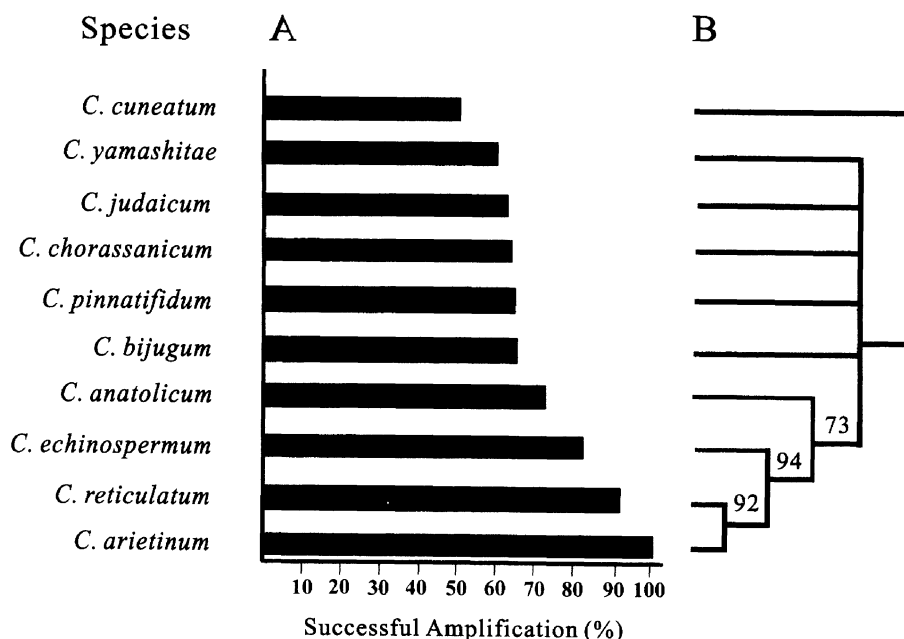


Figure 10: (A) Extent of successful amplification of SSR loci in various *Cicer* species using STMS primer pairs derived from chickpea. (B) Parsimony tree based on zero-allele information and the NTSYS program package (Rohlf 1993). Numbers indicate bootstrap values (obtained after 500 x bootstrapping).

Null-allele information, reflecting mutations in SSR-flanking sequences may be more reliable for phylogenetic studies across and even within species than the actual *size variation* of an SSR locus, because allele size homoplasy - i. e. the fact that loci of similar size do not share the same sequence - has often been encountered (Ortí et al. 1997, Whitton et al. 1997, Peakall et al. 1998; Primmer and Ellegren 1998; see above). Size homoplasy is not only a problem in nuclear SSRs but has also been detected in chloroplast SSRs of wild perennial relatives of soybean, genus *Glycine* (Doyle et al. 1998). Sequence comparison of highly divergent amplification products from the genomes of chickpea and its relatives revealed that variations either are caused by large differences in the number of SSR repeat units (in the case of chickpea's closest relative *C. reticulatum*), or by the amplification of another locus unrelated to the one amplified from chickpea DNA (in the case of the other crossability groups). Sequence information

altogether suggested that STMS markers derived from chickpea may efficiently and reliably be used for synteny studies in chickpea's crossability group including *C. anatolicum*. However, care should be taken when applying chickpea-derived STMS markers to more distant species, because in many cases loci will not be syntenic.

2.6. Microsatellite enrichment strategies for the generation of STMS

Since the isolation of hundreds of SSR-containing clones from small-insert libraries requires considerable time and effort, a number of strategies were developed for the specific enrichment of SSRs in genomic libraries. In an elegant approach, Ostrander et al. (1992) created a small-insert phagemid library in an *E.coli* strain deficient in UTPase (*dut*) and uracil-N-glycosylase (*ung*) genes. Consequently, in this strain dUTP can efficiently be incorporated into DNA instead of dTTP. Single-stranded phagemid DNA was isolated from such a library, second-strand synthesis primed with $[CA]_n$ and $[TG]_n$ primers, and the products transformed into a wild-type *E.coli* strain. In this strain the strong selection against single-stranded, uracil-containing DNA resulted in a library primarily consisting of primer-extended, double-stranded products which contained 50 times more CA-repeats than the starting fragments.

Another enrichment strategy was proposed by Ito et al. (1992). In this procedure, a small insert restriction fragment library is established prior to enrichment. When total supercoiled plasmid preparations from the library are mixed with a biotinylated oligonucleotide such as $[CT]_{13}$, target DNA sequences (i.e. GA:CT double strands) are captured by means of intermolecular triple helix formation. Positive clones are bound to streptavidin-coated magnetic beads, released by alkali, plasmids are purified and used for secondary transformation. The use of this "triplex affinity capture" technique is of course limited to sequence motifs that are capable of triple helix formation *in vitro* (such as GA- and GAA-repeats; see e.g. Milbourne et al. 1997).

Libraries enriched for MSs may also be obtained by simply cloning MP-PCR or RAMPO products. Fisher et al. (1996) employed AMP-PCR to reach nearly 100 % enrichment of SSRs in genomic libraries from *Pinus radiata*. AMP-PCR was performed at high stringency with a single primer containing a $[CT]_6$ motif fused to a highly degenerated 5'-anchor (sequence KKVRVRVRV $[CT]_6$; K=G/T, V=G/C/A, R=G/A), and resulted in polymorphic banding patterns. When PCR products were cloned and sequenced at random, every clone contained an SSR with a minimum length of 6 $[CT]$ units at either end. Probing a library of 1000 clones generated that

way with $[CT]_{16}$ at high stringency yielded 15 clones, of which 13 were unique and contained at least one $[CT]_n$ SSR with $n > 14$. Five of the clones possessed internal SSRs, indicating a clustering of SSRs at certain loci. Using a similar approach, Brachet et al. (1999) found only 2 internal repeats in AMP-PCR products from *Fraxinus excelsior*. However, sufficient polymorphic PCR products resulted from the 5' AMP-PCR itself: cloning showed that flanking SSRs were included in the amplification products. The successful generation of SSR markers from cloned RAMPO fragments has been described for *Camellia japonica* (Ueno et al. 1999).

The most commonly applied enrichment strategies rely on hybridization selection of simple-sequence repeats prior to cloning (Karagyzov et al. 1993; Armour et al. 1994; Kijas et al. 1994; Kandpal et al. 1994; Waldbieser 1995, Edwards et al. 1996, Prochazka 1996, White and Powell 1997a, Connel et al. 1998, Fischer and Bachmann 1998, Koblizkova et al. 1998; Cordeiro et al. 1999). These techniques generally involve the following steps: (1) Fragmentation of genomic DNA either by sonication, nebulization, or by digestion with one or more restriction enzymes, (2) ligation of fragments to adaptors that allow a "whole genome PCR" at this or a later stage of the procedure and (3) amplification, denaturation and hybridization of fragments to SSR-specific oligonucleotides immobilized on a nylon membrane, or bound to magnetic beads via biotin-streptavidin interaction. In this latter case, (4) SSR-containing fragments are separated from unbound fragments by a magnetic device. Non-annealed fragments are washed off and (5) hybridized fragments released by boiling or alkali treatment (6). The fragments enriched for SSRs are reamplified using adaptor-complementary primers, digested with restriction enzymes to remove the adaptors (7), and the product ligated into a vector (8) and transformed into *E.coli*.

Random sequencing generally proved the presence of one or more SSRs in 50-70% of the clones obtained by these procedures. However, when these libraries were probed under high-stringency conditions, enrichment was around 20 % as compared to the 75 % detected by random sequencing (Connell et al. 1998). This indicates that during the enrichment process also sequences with very short or cryptic SSRs are captured that are not detected by high-stringency hybridization. Further, in species with very large genomes such as conifers, enrichment efficiencies below 10% have frequently been encountered. In these cases, post-selection identification of colonies with larger SSRs by hybridization with digoxigenin (DIG)-labeled SSR-complementary oligonucleotides increased the number of useful sequences from 1.5 % to 98% in *Araucaria cunninghamii*, and from 0 to 100% in *Pinus elliottii* (Scott et al. 1999). In sugarcane, initial enrichment by the method of Edwards et al. (1996) was only 20 %. This number could be increased to 50 % as judged from random sequencing by optimizing salt and washing conditions (Cordeiro et al. 1999). Enrichment efficiency could be

further increased by using the Stoffel fragment of *Taq* polymerase for PCR, and either sonication or combinations of different restriction enzymes for the generation of suitably sized DNA fragments (Thompson et al. 1999).

Further improvements of the enrichment process may come from the use of a specifically designed blunt-ended linker called SNX, that contains part of a restriction site which is formed upon dimerisation of the linker. During ligation of the linker to blunt-ended genomic DNA prior to the enrichment with magnetic beads, the inclusion of the respective restriction enzyme in the ligation mix prevents linker-dimers and increases the amount of linkers available for their ligation to the genomic DNA. The use of a blunt-ended linker allows the simultaneous digestion with several restriction enzymes, resulting in fragments <1000 bp on average, which circumvents their size-selection by electrophoresis. After enrichment, the fragments are amplified by PCR and ligated to a vector. Again, a restriction enzyme is included in the reaction to prevent the formation of chimeric fragments. SSR-enriched libraries of tree and animal species produced that way contained only low numbers of redundant clones (Hamilton et al. 1999).

A problem frequently encountered - and especially pronounced in enriched libraries - is the sometimes small proportion of SSR loci that are actually useful for marker generation. For example, of the 831 clones sequenced from a tea tree library enriched according to Edwards et al. (1996), only 11.2 % contained polymorphic SSRs (Rosetto et al. 1999). This discouraging low number originated from the fact that many clones (24.3 %) were sequenced twice, 35.2 % contained only short SSRs (less than 12 dinucleotide or 8 trinucleotide units, respectively), and 28.9 % of SSR-flanking sequences were too short to allow the design of primers. Of the 139 primer pairs finally designed, only 102 (73.4 %) produced a fragment of the expected size, and only 93 (66.9%) detected a polymorphism (Rosetto et al. 1999). Duplicates were also reported in other enrichment studies (Koblizkova et al. 1998, Neu et al. 1999, Becher et al. 2000), and are most certainly a consequence of the PCR step after enrichment. Multiple copies of identical or highly similar SSR loci may also be isolated if the SSR is part of a larger repeat, as was the case in sugarbeet (Mörchen et al. 1996) and lettuce (Van de Wiel et al. 1999).

Another problem is the generation of chimeric clones, i.e. clones in which different genomic regions have been joined together to form a single insert. Chimeras can often be identified by computerized sequence analysis, because the recognition site of the enzyme used for cloning is re-formed at the joining site (except when the library was established from sheared DNA, or from DNA cut by several different enzymes. Koblizkova et al. (1998) described a mechanism of chimera formation, which appears to be specific for the PCR step during microsatellite enrichment procedures. Few residual SSR-specific oligonucleotides that remain in the sample from the hybridization selection step may pair to a SSR-containing genomic fragment,

and act as a primer in the forthcoming PCR. One portion of the target is then amplified together with the adapter primer. In the next round of amplification, the resulting PCR product may again pair to a SSR present in another fragment, and is then extended to the other side. The resulting hybrid molecule contains parts of both original restriction fragments involved, with a common SSR motif in the middle (see Fig. 1 in Koblizkova et al. 1998 for illustration). Such chimeras will normally remain undetected. Koblizkova et al. (1998) eliminated these artifacts by using 3'-modified oligonucleotides for the capturing step that cannot be elongated by DNA polymerase. The resulting library then contained less redundant clones, and all primer pairs amplified correctly sized products from genomic DNA.

We tested several enrichment strategies ourselves, and encountered all the usual problems as e.g. sequencing the same locus several times, or being unable to produce an amplification fragment from the primers derived from SSR-flanking sequences. Nevertheless, we were able to generate multiple STMS markers for the banana pathogen *Mycosphaerella Fijensis* (Neu et al. 1999) and the ornamental plant *Pelargonium hortorum* (Becher et al. 2000), following the method of Fischer and Bachmann (1998). Since adjusting preselection and enrichment conditions may cost considerable time and work, we currently consider the use of enriched vs. unenriched libraries of size-selected DNA fragments equally efficient, at least for some species. For example, for the fungal chickpea pathogen *Ascochyta rabiei* (Geistlinger et al. 2000), which probably has a genome of similar complexity as that of *Mycosphaerella*, and the tropical tree species *Simarouba amara* (Rodriguez et al. 2000), STMS markers were isolated from unenriched, small-insert libraries without problems and in a much shorter time.

2.7. Multiplex PCR and gel electrophoretic detection of STMS polymorphisms: Different approaches

Allele size differences, especially of tri- and tetranucleotide repeats, are often sufficiently large to allow their resolution on high percentage agarose (e.g. Bell and Ecker 1994; Becker and Heun 1995a; Hüttel et al. 1999, see Fig. 9) or nondenaturing polyacrylamide gels in combination with either ethidium bromide (Scrimshaw 1992) or silver staining (Klinkicht and Tautz 1992; Neilan et al. 1994). The resolution power of agarose gels for A/T-rich fragments can be further increased by addition of bis-benzimide polyethylene glycol, enabling the separation of a 302 bp fragment from a 306 bp allele (one tetranucleotide repeat unit) in 2 % NuSieve agarose (Kristensen and Borresen-Dale 1997).

If the appropriate, yet expensive equipment is available, STMS primers can also be labeled with fluorescent dyes, the generated fragments

separated on semi-automatic DNA sequencers, and PCR products detected by real-time laser scanning during gel electrophoresis (Ziegle et al. 1992; Schwengel et al. 1994, Kresovich et al. 1995, Diwan and Cregan 1997, Mitchell et al. 1997). STMS typing using fluorescence has some major advantages over standard techniques involving autoradiography, ethidium bromide or silver staining. First, the use of internal size markers allows the computer to generate a calibration curve for automatic allele sizing and quantitation. This obviates the problem of lane-to-lane and gel-to-gel variation (e.g. band shifts and "smiling" effects). Second, the allelic information is immediately stored and tabulated in a computer as the PCR fragments pass through the detection window. No post-electrophoresis gel handling is required. Third, several STMS markers as well as the size marker (each labelled with a different fluorophore) can be combined on a single lane. This results in high multiplex ratios, especially if differentially labelled primers are also combined into a single PCR assay. However, problems of data acquisition have also been reported. For example, STMS primer pairs from wheat and barley often produced unspecific background bands, precluding multiple sample loading for the same gel run (Donini et al. 1998). This problem could potentially be solved by combining multiplexing with a touch-down PCR protocol (Don et al. 1991), that considerably reduced the number of unspecific bands and allowed the amplification of up to three different loci at the same time (Rithidech et al. 1997). A further problem is that multiplexing PCR reactions with several primer pairs rarely generates products as consistently as single primer reactions do. Sequential loading of samples, though less time-efficient, was the most successful strategy to increase analytical economy in the wheat and barley study (Donini et al. 1998). To overcome these problems, Narvel et al. (2000) developed sets of markers with specifically adjusted primer binding sites to obtain amplified alleles in an optimal size range. A total of 11 multiplex sets containing 74 markers altogether, and covering all the 20 soybean chromosomes were generated that way. The size adjustment of alleles by appropriate primer design, and the use of different fluorescent labels for markers with similar allele sizes allowed to combine up to 8 primer pairs in a single PCR and the simultaneous separation of markers in a single lane.

3. PERSPECTIVES

Compared to few years ago, the number of molecular marker techniques and the applicability of the various dominant and codominant DNA markers have facilitated genome analysis in a way that molecular breeding became a reality. One emerging elite marker type, the so-called single nucleotide polymorphisms (SNPs), probably are the most advanced

markers to date. SNPs comprise the largest set of sequence variants in most organisms, and encircle both single nucleotide polymorphisms (transitions, transversions) and small insertions and deletions. For example, genome-wide SNP maps have been constructed for the yeast *Saccharomyces cerevisiae* (Winzeler et al. 1998) and the plant *Arabidopsis thaliana* (Cho et al. 1999), but are also available for other eukaryotes (e.g. human and mouse; Wang et al. 1998). The resolution of these maps varies from 1.0 to 3.5 cM, depending on the organism. SNPs will attract much more attention in future, since specific SNPs in particular gene or promoter sequences are supposed to be associated (if not responsible) for a series of diseases in humans and animals. Therefore, such SNPs serve as diagnostic tools in medicine. Also, SNPs as products of high-throughput genomic sequencing procedures (Hoogendorn et al. 1999) can be used to saturate existing genetic and physical maps.

Microsatellite markers played, and will play a comparably important role as the SNPs. We expect that in future closely linked flanking microsatellite markers will be available for every useful trait, at least in important crops. Multiplexing, high-throughput screening and computer-based data management all have been proven to work with STMS markers as well. Therefore any desired genotype will be defined by unique microsatellite bands detected on sequencing gels by fluorescence laser scanning very soon. Data can be stored and processed by computers, which then identify useful offspring by their expected banding patterns or allele sizes. The assignment of such a "bar-code" to each and every genotype will help to protect breeders rights, distinguish offspring of "real" crosses from "false" F1s derived from selfing, and clarify paternity in outcrossing species. Last but not least, STMS markers by their virtue represent ideal anchors for YAC- or BAC-based cloning of agronomically relevant genes, that then can be used to complement the genetic make-up of plants.

Acknowledgements

This article is dedicated to all our colleagues who contributed substantially to legume research in the Plant Molecular Biology Laboratory, including M. Anis (India), R. Arreguin-Espinoza (Mexico), A.M. Benko-Iseppon, D. P. Biradar (India), L. Brennscheidt, T. Bünger, W. Choumane (Syria), G. Gortner, C. Jacobi, R. Jungmann, M. Klatt, J. Lichtenzweig (Israel), F. J. Muehlbauer (USA), J. Juarez Muñoz (Mexico), Y. Duran Vinagre (Spain), T. Pfaff, J. Ramser, S. Sahi (Pakistan), P. C. Sharma (India), G. Sonnante (Italy), T. Staginnus, S. M. Udupa (India), and H.T. Wolf. Research of the authors is supported by grants from the German Ministry of Technical Cooperation (BMZ grant 89.7860.3-01.130), the German Research Council (DFG grants Ka 332/17-1, -17-2, and 17 -3), and the International Atomic Energy Agency (IAEA grant 302-D2-GFR-10974).

Acronyms	
AFLP	<i>Amplified Fragment Length Polymorphism</i>
AMP-PCR	<i>Anchored Microsatellite-primed PCR</i>
BAC	<i>Bacterial Artificial Chromosome</i>
BSA	<i>Bulked Segregant Analysis</i>
DAF	<i>DNA Amplification Fingerprinting</i>
dRAMP	<i>Digested Random Amplified Microsatellite Polymorphism</i>
EST	<i>Expressed Sequence Tag</i>
IRAP	<i>Inter-Retrotransposon Amplified Polymorphism</i>
IRS	<i>Interspersed Repetitive Sequence</i>
ISSR	<i>Inter-Simple Sequence Repeat</i>
LTR	<i>Long Terminal Repeat</i>
MAS	<i>Marker-assisted Selection</i>
MP-PCR	<i>Microsatellite-primed PCR</i>
MS	<i>Microsatellite</i>
PCR	<i>Polymerase Chain Reaction</i>
RAHM	<i>Random Amplified Hybridization Microsatellites</i>
RAMP	<i>Random Amplified Microsatellite Polymorphism</i>
RAMPO	<i>Random Amplified Microsatellite Polymorphism</i>
RAMS	<i>Random Amplified Microsatellites</i>
RAPD	<i>Random Amplified Polymorphic DNA</i>
REMAP	<i>Retrotransposon-Microsatellite Amplified Polymorphism</i>
RFLP	<i>Restriction Fragment Length Polymorphism</i>
SAMPL	<i>Selective Amplification of Polymorphic Loci</i>
SPAR	<i>Single Primer Amplification Reaction</i>
SSLP	<i>Simple Sequence Length Polymorphism</i>
SSR	<i>Simple Sequence Repeat</i>
STMS	<i>Sequence-tagged Microsatellite Site</i>
STR	<i>Short Tandem Repeat</i>
UTR	<i>Untranslated Region</i>
VNTR	<i>Variable Number of Tandem Repeat</i>
YAC	<i>Yeast Artificial Chromosome</i>

References

- Ahmad, F. and Slinkard, A.E. (1992) Genetic relationships in the genus *Cicer* L as revealed by polyacrylamide gel electrophoresis of seed storage proteins. *Theor Appl Genet* **84**, 688-692.
- Akagi, H. Yokozeki, Y. Inagaki, A. and Fujimura, T. (1997) Highly polymorphic microsatellites of rice consist of AT repeats, and classification of closely related cultivars with these microsatellite loci. *Theor Appl Genet* **94**, 61-67.
- Akkaya, M.S., Bhagwat, A.A. and Cregan, P.B. (1992). Length polymorphisms of simple sequence repeat DNA in soybean. *Genetics* **132**, 1131-1139.
- Akkaya, M.S., Shoemaker, R.C., Specht, J.E., Bhagwat, A.A. and Cregan, P.B. (1995). Integration of simple sequence repeat DNA markers into a soybean linkage map. *Crop Sci* **35**, 1439-1445.
- Ali, S. Müller C.R. and Epplen, J.T. (1986) DNA fingerprinting by oligonucleotide probes specific for simple repeats. *Hum Genet* **74**, 239-243.
- Arens, P., Odinet, P., Van Heusden, A.W., Lindhout, P. and Vosman, B. (1995) GATA- and GACA-repeats are not evenly distributed throughout the tomato genome. *Genome* **38**, 84-90.
- Arshchenkova, T. and Ganai, M.W. (1999) Long tomato microsatellites are predominantly associated with centromeric regions. *Genome* **42**, 536-544.
- Armour, J.A.L., Neumann, R., Gobert, S. and Jeffreys, A.J. (1994) Isolation of human simple repeat loci by hybridization selection. *Hum. Mol. Genet.* **3**, 599-605.
- Becher, S.A., Steinmetz, K., Weising, K., Boury, S., Peltier, D., Renou, J.P., Kahl, G. and Wolff, K. (2000) Microsatellites for cultivar identification in *Pelargonium*. *Theor. Appl. Genet.*, in press.
- Becker, J. and Heun, M. (1995a) Barley microsatellites: allele variation and mapping. *Plant Mol Biol* **27** 835-845.
- Becker, J. and Heun, M. (1995b) Mapping of digested and undigested random amplified microsatellite polymorphisms in barley. *Genome* **38**, 991-998.
- Beckmann, J.S. and Soller, M. (1990) Toward a unified approach to genetic mapping of eukaryotes based on sequence tagged microsatellite sites. *Bio/Technology* **8**, 930-932.
- Bell, C.J. and Ecker, J.R. (1994) Assignment of 30 microsatellite loci to the linkage map of *Arabidopsis*. *Genomics* **19**, 137-144.
- Botstein, D., White, R.L., Skolnik, M. and Davis, R.W. (1980) Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *Am. J. Hum. Genet.* **32**, 314-331.
- Brachet, S., Jubier, M.F., Richard, M., Jung-Muller, B. and Frascaria-Lacoste, N. (1999) Rapid identification of microsatellite loci using 5' anchored PCR in the common ash *Fraxinus excelsior*. *Mol. Ecol.* **8**, 160-163.

- Brandes, A. Thompson, H., Dean, C. and Heslop-Harrison, J.S. (1997) Multiple repetitive sequences in the paracentromeric regions of *Arabidopsis thaliana* L. Chromosome Res. **5**, 238-246.
- Brinkmann, B., Klintschar, M., Neuhuber, F., Hühne, J. and Rolf, B. (1998) Mutation rate in human microsatellites: influence of the structure and length of the tandem repeat. Am. J. Hum. Genet. **62**, 1408-1415.
- Broun, P. and Tanksley, S.D. (1993) Characterization of tomato clones with sequence similarity to human minisatellites 33.6 and 33.15. Plant Mol. Biol. **23**, 231-242.
- Broun, P. and Tanksley, S.D. (1996) Characterization and genetic mapping of simple repeat sequences in the tomato genome. Mol. Gen. Genet. **250**, 39-49.
- Bryan, G.J., Collins, A.J., Stephenson, P., Orry, A., Smith, J.B. and Gale, M.D. (1997) Isolation and characterization of microsatellites from hexaploid bread wheat. Theor. Appl. Genet. **94**, 557-563.
- Bryan, G.J., McNicoll, J., Ramsey, G., Meyer, R.C. and De Jong, W.S. (1999) Polymorphic simple sequence repeat markers in chloroplast genomes of Solanaceous plants. Theor. Appl. Genet. **99**, 859-867.
- Caetano-Anollés, G., Bassam, B.J. and Gresshoff, P.M. (1991) DNA amplification fingerprinting: a strategy for genome analysis. Plant Mol. Biol. Rep. **9**, 292-305.
- Chakraborty, R., Kimmel, M., Strivers, D.N., Davison, L.J. and Deka, R. (1997) Relative mutation rates at di- tri- and tetranucleotide microsatellite loci. Proc. Natl. Acad. Sci. USA **94**, 1041-1046.
- Chase, M., Kesseli, R. and Bawa, K. (1996) Microsatellite markers for population and conservation genetics of tropical trees. Am. J. Bot. **83**, 51-57.
- Chen, X.S., Temnykh, Y., Xu, Y., Cho, Y.G. and McCouch, S.R. (1997) Development of a microsatellite map providing genome-wide coverage in rice (*Oryza sativa* L.) Theor. Appl. Genet. **95**, 553-567.
- Chin, E.C.L., Senior, M.L., Shu, H. and Smith, J.S.C. (1996) Maize simple repetitive DNA sequences: abundance and allele variation. Genome **39**, 866-873.
- Cho, R.J. et al. (1999) Genome-wide mapping with biallelic markers in *Arabidopsis thaliana*. Nature Genetics **23**, 203-207.
- Cho, Y.G., Ishii, T., Temnykh, S., Chen, X., Lipovich, L., McCouch, S.R., Park, W.D., Ayres, N. and Cartinhour, S. (2000) Diversity of microsatellites derived from genomic libraries and GeneBank sequences in rice (*Oryza sativa* L.). Theor. Appl. Genet. **100**, 713-722.
- Choumane, W., Winter, P., Weigand, F. and Kahl, G. (2000) Conservation and variability of sequence tagged microsatellite sites from chickpea (*Cicer arietinum* L.) within the genus *Cicer*. Theor. Appl. Genet. **101**, 269-278.
- Cifarelli, R.A., Gallitelli, M. and Cellini, F. (1995) Random amplified hybridization microsatellites (RAHM): isolation of a new class of microsatellite-containing DNA clones. Nucleic Acids Res. **23**, 3802-3803.

- Cipriani, G., Lot, G., Huang, W.-G., Marrazzo, M.T., Peterlunger, E. and Testoli, R. (1999) AC/GT and AG/CT microsatellite repeats in peach [*Prunus persica* (L) Batsch]: isolation characterization and cross-species amplification in *Prunus*. *Theor. Appl. Genet.* **99**, 65-72.
- Connel, J.P., Pammi, S., Iqbal, M.J., Huizinga, T. and Reddy, A.S. (1998) A high-throughput procedure for capturing microsatellites from complex genomes. *Plant Mol. Biol. Rep.* **16**, 341-349.
- Cordeiro, G.M., Maguire, T.L., Edwards, K.J. and Henry, R.J. (1999) Optimisation of a microsatellite enrichment technique in *Saccharum* spp. *Plant Mol. Biol. Rep.* **17**, 225-229.
- Cregan, P.B., Jarvik, T., Bush, A.L., Shoemaker, R.C., Lark, K.G., Kahler, A.I., Kaya, N., VanThoi, T.T., Lohnes, D.G., Chung, J. and Specht, J. (1999a) An integrated genetic linkage map of the soybean genome. *Crop Sci.* **39**, 211-217.
- Cregan, P.B., Mudge, J., Fickus, E.W., Marek, L.F., Danesh, D., Denny, R., Shoemaker, R.C., Mathews, B.F., Jarvik, T. and Young, N.D. (1999b) Targeted isolation of simple sequence repeat markers through the use of bacterial artificial chromosomes. *Theor. Appl. Genet.* **98**, 919-928.
- Dávila, J.A., Loarce, Y. and Ferrer, E. (1999) Molecular characterization and genetic mapping of random amplified microsatellite polymorphism in barley. *Theor. Appl. Genet.* **98**, 265-273.
- Dayanandan, S., Bawa, K.S. and Kesseli, R. (1997) Conservation of microsatellites among tropical trees. *Am. J. Bot.* **84**, 1658-1663.
- Deka, R., Shriver, M.D., Yu, L.M., Aston, C.E., Chakraborty, R. and Ferrell, R.E. (1994) Conservation of human chromosome 13 polymorphic microsatellite (CA)_n repeats in chimpanzees. *Genomics* **22**, 226-230.
- Delseny, M., Cooke, R., Raynal, M. and Grellet, F. (1997) The *Arabidopsis thaliana* cDNA sequencing projects. *FEBS Lett.* **405**, 129-132.
- Depeiges, A., Goubely, C., Lenoir, A., Cocherel, S., Picard, G., Raynal, M., Grellet, F. and Delseny, M. (1995) Identification of the most represented repeated motifs in *Arabidopsis thaliana* microsatellite loci. *Theor. Appl. Genet.* **91**, 160-168.
- Dib, C., Fauré, S., Fizames, C., Samson, D., Drouot, N., Vignal, A., Millasseau, P., Marc, S., Hazan, J. Seboun, E., Lathrop, M., Gyapay, G., Morissette, J. and Weissenbach, J. (1996) A comprehensive genetic map of the human genome based on 5,264 microsatellites. *Nature* **380**, 152-154.
- Dietrich, W.F., Miller, J., Steen, R., Merchant, M.A., Damron-Boles, D., Husain, Z., Dredge, R., Daly, M., Ingalls, K.A., O'Connor, T.J., Evans, C.A., DeAngelis, M.M., Levinson, D.M., Kruglyak, L., Goodman, N., Copeland, N.G., Jenkins, N.A., Hawkins, T.L., Stein, L., Page, D.C. and Lander, E.S. (1996) A comprehensive genetic map of the mouse genome. *Nature* **380**, 149-152.
- Dirlewanger, E., Isaac, P.G., Ranade, S., Belajouza, M., Cousin, R. and De Vienne, D. (1994) Restriction fragment length polymorphism analysis of loci associated with disease resistance genes and developmental traits in *Pisum sativum* L. *Theor. Appl. Genet.* **88**, 17-27.

- Diwan, N. and Cregan, P.B. (1997) Automated sizing of fluorescent-labelled simple sequence repeat (SSR) markers to assay genetic variation in soybean. *Theor. Appl. Genet.* **95**, 723-733.
- Diwan, N., Bhagwat, A.A., Bauchan, G.B. and Cregan, P.B. (1997) Simple sequence repeat DNA markers in alfalfa and perennial and annual *Medicago* species. *Genome* **40**, 887-895.
- Don, R.H., Cox, P.T., Wainwright, K., Baker, K. and Mattock, J.S. (1991) "Touchdown" PCR to circumvent spurious priming during gene amplification. *Nucl Acids Res.* **19**, 4008.
- Donini, P., Stephenson, P., Bryan, G.J. and Koebner, R.M.D. (1998) The potential of microsatellites for high-throughput genetic diversity assessment in wheat and barley. *Genet Resources Crop. Evol.* **45**, 415-421.
- Doyle, J.J., Morgante, M., Tingey, S.V. and Powell, W. (1998) Size homoplasmy in chloroplast microsatellites of wild perennial relatives of soybean (*Glycine* subgenus *Glycine*). *Mol. Biol. Evol.* **15**, 215-218.
- Echt, C.S. and May-Marquardt, P. (1997) Survey of microsatellite DNA in pine. *Genome* **40**, 9-17.
- Echt, C.S., DeVerno, L.L., Anzidei, M. and Vendramin, G.G. (1998) Chloroplast microsatellites reveal population genetic diversity in red pine, *Pinus resinosa* Ait. *Mol. Ecol.* **7**, 307-316.
- Edwards, A., Civitello, A., Hammond, H.A. and Caskey, C.T. (1991) DNA typing and genetic mapping with trimeric and tetrameric tandem repeats. *Am. J. Hum. Genet.* **49**, 746-756.
- Edwards, K.J., Barker, J.H.A., Daly, A., Jones, C. and Karp, A. (1996) Microsatellite libraries enriched for several microsatellite sequences in plants. *BioTechniques* **20**, 758-760.
- Ender, A., Schwenk, K., Städler, T., Streit, B. and Schierwater, B. (1996) RAPD identification of microsatellites in *Daphnia*. *Mol. Ecol.* **5**, 437-441.
- Epplen, J.T. (1992) The methodology of multilocus DNA fingerprinting using radioactive or nonradioactive probes specific for simple repeat motifs. In: Chrambach, A. Dunn M, Radola BJ (eds.), *Advances in Electrophoresis*, Vol.5. VCH, Weinheim, Germany, pp 59-112.
- Epplen, J.T., Mäueler, W. and Santos, E.J.M. (1998) On GATAGATA and other "junk" in the barren stretch of genomic desert. *Cytogenet. Cell Genet.* **80**, 75-82.
- Fischer, D. and Bachmann, K. (1998) Microsatellite enrichment in organisms with large genomes (*Allium cepa* L.) *Biotechniques* **24**, 796-798.
- Fisher, P.J., Gardner, R.C. and Richardson, T.E. (1996) Single locus microsatellites isolated using 5' anchored PCR. *Nucl Acids Res.* **24**, 4369-4372.
- FitzSimmons, N.N., Moritz, C. and Moore, S.S. (1995) Conservation and dynamics of microsatellite loci over 300 million years of marine turtle evolution. *Mol. Biol. Evol.* **12**, 432-440.
- Forbes, S.H., Hogg, J.T., Buchanan, F.C., Crawford, A.M. and Allendorf, F.W. (1995) Microsatellite evolution in congeneric mammals: domestic and bighorn sheep. *Mol. Biol. Evol.* **12**, 1106-1113.

- Garza, J.C., Slatkin, M. and Freimer, N.B. (1995) Microsatellite allele frequencies in humans and chimpanzees, with implications for constraints on allele size. *Mol. Biol. Evol.* **12**, 594-603.
- Gaur, P.M. and Slinkard, A.E. (1990) Genetic control and linkage relations of additional isozyme markers in chickpea. *Theor. Appl. Genet.* **80**, 648-656.
- Geistlinger, J., Weising, K., Winter, P., and Kahl, G. (2000) Locus-specific microsatellite markers for the fungal chickpea pathogen *Didymella rabiei* (anamorph) *Ascochyta rabiei*. *Mol. Ecol.* **9**, 1939-1941
- Gortner, G., Nenno, M., Weising, K., Zink, D., Nagl, W. and Kahl, G. (1998) Chromosomal localization and distribution of simple sequence repeats and the *Arabidopsis*-type telomere sequence in the genome of *Cicer arietinum* L. *Chrom. Res.* **6**, 97-104.
- Gupta, M., Chyi, Y.-S., Romero-Severson, J. and Owen, J.L. (1994) Amplification of DNA markers from evolutionary diverse genomes using single primers of simple-sequence repeats. *Theor. Appl. Genet.* **89**, 998-1006.
- Gupta, P.K. and Varshney, R.K. (2000) The development and use of microsatellite markers for genetic analysis and plant breeding with emphasis on bread wheat. *Euphytica* **113**, 163-185.
- Hamada, H., Seidman, M., Howard, B.H. and Gorman, C.M. (1984) Enhanced gene expression by the poly (dT-dG poly (dC-dA) sequence. *Mol. Cell. Biol.* **4**, 2622-2630.
- Hamilton, M.B., Pincus, E.L., Di Fiore, A. and Fleischer, R.C. (1999) Universal linker and ligation procedures for construction of genomic DNA libraries enriched for microsatellites. *BioTechniques* **27**, 500-507.
- Harr, B., Zangerl, B., Brem, G. and Schlötterer, C. (1998) Conservation of locus-specific microsatellite variability across species: a comparison of two *Drosophila* sibling species, *D. melanogaster* and *D. simulans*. *Mol. Biol. Evol.* **15**, 176-184.
- Hauge, X.Y. and Litt, M. (1993) A study of the origin of "shadow bands" seen when typing dinucleotide repeat polymorphisms by the PCR. *Hum. Mol. Genet.* **2**, 411-415.
- Hearne, C.M., McAleer, M.A., Love, J.M., Aitman, J., Cornall, R.J., Gosh, S., Knight, A.M., Prins, J.-B. and Todd, J.A. (1991) Additional microsatellite markers for mouse genome mapping. *Mamm. Genome* **1**, 273-282.
- Hoogendorn, B. et al. (1999) Genotyping single polymorphisms by primer extension and high performance liquid chromatography. *Hum. Genet.* **104**, 89-93.
- Huang, W.-G., Cipriana, G., Morgante, M. and Testolin, R. (1998) Microsatellite DNA in *Actinidia chinensis*: isolation, characterization and homology in related species. *Theor. Appl. Genet.* **97** 1269-1278.
- Huang, J.C. and Sun, M. (2000) Genetic diversity and relationships of sweetpotato and its wild relatives in *Ipomoea* series *Batatas* (*Convolvulaceae*) as revealed by inter-simple sequence repeat (ISSR) and restriction analysis of chloroplast DNA. *Theor. Appl. Genet.* **100**, 1050-1060.

- Hüttel, B. (1996) Mikrosatelliten als molekulare Marker in der Kichererbse (*Cicer arietinum* L.) Ph.D. Thesis, University of Frankfurt, Germany
- Hüttel, B., Winter, P., Weising, K., Choumane, W., Weigand, F. and Kahl, G. (1999) Sequence-tagged microsatellite site markers for chickpea (*Cicer arietinum* L.). *Genome* **42**, 210–217.
- Ishii, T. and McCouch, S.R. (2000) Microsatellites and microsynteny in the chloroplast genomes of *Oryza* and eight other Gramineae species. *Theor. Appl. Genet.* **100**, 1257–1266.
- Ito, T., Smith, C.L. and Cantor, C.R. (1992) Sequence-specific DNA purification by triplex affinity capture. *Proc. Natl. Acad. Sci. USA* **89**, 495–498.
- Jacob, H.J., Brown, D.M., Bunker, R.K., Daly, M.J., Dzau, V.J., Goodman, A., Koike, G., Kren, V., Kurtz, T., Lernmark, A., Levan, G., Mao, Y., Pettersson, A., Pravenec, M., Simon, J.S., Szpirer, C., Szpirer, J., Trolliet, M.R., Winer, E.S. and Lander, E.S. (1995) A genetic linkage map of the laboratory rat, *Rattus norvegicus*. *Nature Genet.* **9**, 63–69.
- Jin, L., Macaubas, C., Hallmayer, J., Kimura, A. and Mignot, E. (1996) Mutation rate varies among alleles at a microsatellite locus: phylogenetic evidence. *Proc. Natl. Acad. Sci. USA* **93**, 15285–15288.
- Kaemmer, D., Afza, R., Weising, K., Kahl, G. and Novak, F.J. (1992) Oligonucleotide and amplification fingerprinting of wild species and cultivars of banana (*Musa* spp.). *Bio/Technology* **10**, 1030–1035.
- Kaemmer, D., Fischer, D., Jarret, R.L., Baurens, F.C., Grapin, A., Dambier, D., Noyer, J.L., Lannaud, C., Kahl, G., and Lagoda, P.J.L. (1997) Molecular breeding in the genus *Musa*: a strong case for STMS marker technology. *Euphytica* **96**, 49–63.
- Kalendar, R., Grob, T., Regina, M., Suoniemi, A. and Schulman, A. (1999) IRAP and REMAP: two new retrotransposon-based DNA fingerprinting techniques. *Theor. Appl. Genet.* **98**, 705–711.
- Kandpal, R.P., Kandpal, G. and Weissman, S.M. (1994) Construction of libraries enriched for sequence repeats and jumping clones, and hybridization selection for region-specific markers. *Proc. Natl. Acad. Sci. USA* **91**, 88–92.
- Kantety, R.V., Zeng, X., Bennetzen, J.L. and Zehr, B.E. (1995) Assessment of genetic diversity in dent and popcorn (*Zea mays* L.) inbred lines using inter-simple sequence repeat (ISSR) amplification. *Mol. Breeding* **1**, 365–373.
- Karagoyozov, L., Kalcheva, I.D. and Chapman, V.M. (1993) Construction of random small-insert genomic libraries highly enriched for simple sequence repeats. *Nucleic Acids Res.* **21**, 3911–3912.
- Karhu, A., Dieterich, J.H. and Savolainen, O. (2000) Rapid expansion of microsatellite sequences in pines. *Mol. Biol. Evol.* **17**, 259–265.
- Kashi, Y., Nave, A., Darvasi, A., Gruenbaum, Y., Soller, M. and Beckmann, J.S. (1994) How is it that microsatellites and random oligonucleotides uncover DNA fingerprint patterns? *Mamm. Genome* **5**, 525–530.

- Kashi, Y., King, D. and Soller, M. (1997) Simple sequence repeats as a source of quantitative genetic variation. *Trends Genet.* **13**, 74-78.
- Kazan, K. and Muehlbauer, F.J. (1991) Allozyme variation and phylogeny in annual species of *Cicer* (*Leguminosae*). *Pl. Syst. Evol.* **175**, 11-21.
- Kazan, K., Muehlbauer, F.J., Weeden, N.F. and Ladizinsky, G. (1993) Inheritance and linkage relationships of morphological and isozyme loci in chickpea (*Cicer arietinum* L.). *Theor. Appl. Genet.* **86**, 417-426.
- Kijas, J.M.H., Fowler, J.C.S., Garbett, C.A. and Thomas, M.R. (1994) Enrichment of microsatellites from the *Citrus* genome using biotinylated oligonucleotide sequences bound to streptavidin-coated magnetic particles. *BioTechniques* **16**, 658-662.
- Kijas, J.M.H., Fowler, J.C.S. and Thomas, M.R. (1995). An evaluation of sequence tagged microsatellite site markers for genetic analysis within *Citrus* and related species. *Genome* **38**, 349-355.
- Kijas, J.M.H., Thomas, M.R., Fowler, J.C.S. and Roose, M.L. (1997) Integration of trinucleotide microsatellites into a linkage map of *Citrus*. *Theor. Appl. Genet.* **94**, 701-706.
- Klinkicht, M. and Tautz, D. (1992) Detection of simple sequence length polymorphisms by silver staining. *Mol. Ecol.* **1**, 133-134.
- Koblizkova, A., Dolezel, J. and Macas, J. (1998) Subtraction with 3'-modified oligonucleotides eliminates amplification artifacts in libraries enriched for microsatellites. *BioTechniques* **25**, 32-38.
- Kojima, T., Nagaoka, T., Noda, K. and Ogihara, Y. (1998) Genetic linkage map of ISSR and RAPD markers in Einkorn wheat in relation to that of RFLP markers. *Theor. Appl. Genet.* **96**, 37-45.
- Kondo, Y., Sato, K., Kitada, K. and Serikawa, T. (1997) Application of mouse microsatellite markers to rat genome mapping. *Transplantation Proc.* **29**, 1766-1767.
- Korzun, V., Röder, M.S., Wendehake, K., Pasqualone, A., Lotti, C., Ganai, M.W. and Blanco, A. (1999) Integration of dinucleotide microsatellites from hexaploid bread wheat into a genetic linkage map of durum wheat. *Theor. Appl. Genet.* **98**, 1202-1207.
- Kresovich, S., Szewc-McFadden, A.K., Bliks, S.M. and McFerson, J.R. (1995) Abundance and characterization of simple-sequence repeats (SSRs) isolated from a size-fractionated genomic library of *Brassica napus* L. (rapeseed). *Theor. Appl. Genet.* **91**, 206-211.
- Kristensen, V.N. and Borresen-Dale, A.-L. (1997) Improved electrophoretic separation of polymorphic short tandem repeats in agarose gels using bis-Benzimide. *BioTechniques* **23**, 634-636.
- Labdi, M., Robertson, L.D., Singh, K.B. and Charrier, A. (1996) Genetic diversity and phylogenetic relationships among the annual *Cicer* species as revealed by isozyme polymorphisms. *Euphytica* **88**, 181-188.

- Lagercrantz, U., Ellegren, H. and Andersson, L. (1993). The abundance of various polymorphic microsatellite motifs differs between plants and vertebrates. *Nucleic Acids Res.* **21**, 1111-1115.
- Lench, N.J., Norris, A., Bailey, A., Booth, A. and Markham, A.F. (1996) Vectorette PCR isolation of microsatellite repeat sequences using anchored dinucleotide repeat primers. *Nucleic Acids Res.* **24**, 2190-2192.
- Levinson, G. and Gutman, G.A. (1987) Slipped-strand mispairing: a major mechanism for DNA sequence evolution. *Mol. Biol. Evol.* **4**, 203-221.
- Li, Y.-C., Röder, M.S., Fahima, T., Kirzhner, V.M., Beiles, A., Korol, A.B. and Nevo, E. (2000) Natural selection causing MS diversity in wild emmer wheat at the ecologically variable microsite Ammiad, Israel. *Theor. Appl. Genet.* **100**, 985-999.
- Liu, Z.W., Biyashev, R.M. and Saghai-Marooif, M.A. (1996) Development of simple sequence repeat DNA markers and their integration into a barley linkage map. *Theor. Appl. Genet.* **93**, 869-876.
- Litt, M. and Luty, J.A. (1989) A hypervariable microsatellite revealed by in vitro amplification of a dinucleotide repeat within the cardiac muscle actin gene. *Am. J. Hum. Genet.* **44**, 397-401.
- Lorieux, M., Ndjondjop, M.N. and Ghesquière, A. (2000) A first interspecific *Oryza sativa* X *Oryza glaberrima* microsatellite-based genetic linkage map. *Theor. Appl. Genet.* **100**, 593-601.
- Lu, J., Knox, M.R., Ambrose, M.J., Brown, J.K.M. and Ellis, T.H.N. (1996) Comparative analysis of genetic diversity in pea assessed by RFLP- and PCR-based methods. *Theor. Appl. Genet.* **93**, 1103-1111.
- Manninen, I. and Schulmann, A.H. (1993) BARE1, a copia-like retroelement in barley (*Hordeum vulgare* L.). *Plant Mol. Biol.* **22**, 829-846.
- Mäueler, W., Bassili, G., Arnold, R., Renkawitz, R. and Epplen, J.T. (1999) The (gt)n(ga)m containing intron 2 of HLA-DRB alleles binds a zinc-dependent protein and forms non B-DNA structures. *Gene* **226**, 9-23.
- Marek, L.F. and Shoemaker, R.C. (1997) BAC contig development by fingerprint analysis in soybean. *Genome* **40**, 429-427.
- McCouch, S.R., Chen, X., Panaud, O., Temnykh, S., Xu, Y., Cho, Y.G., Huang, N., Ishii, T. and Blair, M. (1997) Microsatellite marker development, mapping and application in rice breeding. *Plant Mol. Biol.* **35**, 89-99.
- McMurray CT (1995) Mechanisms of DNA expansion. *Chromosoma* **104**, 2-13.
- Meloni, R., Albanese, V., Ravassard, P., Treilhou, F. and Mallet, J. (1998) A tetranucleotide polymorphic microsatellite, located in the first intron of the tyrosine hydroxylase gene, acts as a transcription regulatory element in vitro. *Hum. Mol. Genet.* **7**, 423-428.

- Meyer, W., Mitchell, T.G., Freedman, E.Z. and Vilgalys, R. (1993) Hybridization probes for conventional DNA fingerprinting used as single primers in the polymerase chain reaction to distinguish strains of *Cryptococcus neoformans*. *J. Clin. Microbiol.* **31**, 2274-2280.
- Michelmore, R.W., Paran, I. and Kesseli, R.V. (1991) Identification of markers linked to disease resistance genes by bulked segregant analysis: a rapid method to detect markers in specific genomic regions using segregating populations. *Proc. Natl. Acad. Sci. USA* **88**, 9828-9832.
- Milbourne, D., Meyer, R.C., Collins, A.J., Ramsey, L.D., Gebhardt, C. and Waugh, R. (1998) Isolation, characterization and mapping of simple sequence repeat loci in potato. *Mol. Gen. Genet.* **259**, 233-245.
- Miller, J.C. and Tanksley, S.D. (1990) RFLP analysis of phylogenetic relationships and genetic variation in the genus *Lycopersicon*. *Theor. Appl. Genet.* **80**, 437-448.
- Mitchell, S.E., Kresovich, S., Jester, C.A., Hernandez, C.J. and Szewcz-McFadden, A.K. (1997) Application of multiplex PCR and fluorescence-based, semi-automated allele sizing technology for genotyping plant resources. *Crop Sci.* **37**, 617-624.
- Mörchen, M., Cuguen, J., Michaelis, G., Hänni, C. and Saumitou-Laprade, P. (1996) Abundance and length polymorphism of microsatellite repeats in *Beta vulgaris* L. *Theor. Appl. Genet.* **92**, 326-333.
- Moore, S.S., Sargeant, L.L., King, T.J., Mattick, J.S., Georges, M. and Hetzel, D.J.S. (1991) The conservation of dinucleotide microsatellites among mammalian genomes allows the use of heterologous PCR primer pairs in closely related species. *Genomics* **10**, 654-660.
- Morgante, M. and Olivieri, A.M. (1993) PCR-amplified microsatellites as markers in plant genetics. *Plant J.* **3**, 175-182.
- Morgante, M., Rafalski, A., Biddle, P., Tingey, S. and Olivieri, A.M. (1994) Genetic mapping and variability of seven soybean simple sequence repeat loci. *Genome* **37**, 763-769.
- Nadir, E., Margalit, H., Gallily, T. and Ben-Sasson, S.A. (1996) Microsatellite spreading in the human genome: Evolutionary mechanisms and structural implications. *Proc. Natl. Acad. Sci. USA* **93**, 6470-6475.
- Nakamura, Y., Leppert, M., O'Connell, P., Wolff, R., Holm, T., Culver, M., Martin, C., Fujimoto, E., Hoff, M., Kumlin, E. and White, R. (1987) Variable number of tandem repeat (VNTR) markers for human gene mapping. *Science* **235**, 516-522.
- Nanda, I., Feichtinger, W., Schmid, W., Schröder, J.H., Zischler, H. and Epplen, J.T. (1990) Simple repetitive sequences are associated with the differentiation of the sex chromosomes in the guppy fish. *J. Mol. Evol.* **30**, 456-462.
- Narvel, J.M., Chu, W.-C., Fehr, W.R., Cregan, P.B. and Shoemaker, R.C. (2000) Development of multiple sets of simple sequence repeat DNA markers covering the soybean genome. *Mol. Breeding* **6**, 175-183.
- Neilan, B.A., Leigh, D.A., Rapley, E. and McDonald, B.L. (1994) Microsatellite genome screening: rapid non-denaturing, non-isotopic dinucleotide repeat analysis. *BioTechniques* **17**, 708-712.

- Neu, C., Kaemmer, D., Kahl, G., Fischer, D. and Weising, K. (1999) Polymorphic microsatellite markers for the banana pathogen *Mycosphaerella fijiensis*. *Mol. Ecol.* **8**, 513-525.
- Orti, G., Pearse, D.E. and Avise, J.C. (1997) Phylogenetic assessment of length variation at a microsatellite locus. *Proc. Natl. Acad. Sci. USA* **94**, 10745-10749.
- Ostrander, E.A., Jong, P.M., Rine, J. and Duyk, G. (1992) Construction of small-insert genomic DNA libraries highly enriched for microsatellite repeat sequences. *Proc. Natl. Acad. Sci. USA* **89**, 3419-3423.
- Paglia, G.P., Oliveri, A.M. and Morgante, M. (1998) Towards second-generation STS (sequence-tagged sites) linkage maps in conifers: a genetic map of Norway spruce (*Picea abies* K.). *Mol. Gen. Genet.* **258**, 466-478.
- Panaud, O., Chen, X. and McCouch, S.R. (1995) Frequency of microsatellite sequences in rice (*Oryza sativa* L.). *Genome* **38**, 1170-1176.
- Parasnis, A.S., Ramakrishna, W., Chowdari, K.V., Gupta, V.S. and Ranjekar, P.K. (1999) Microsatellite (GATA)_n reveals sex-specific differences in Papaya. *Theor. Appl. Genet.* **99**, 1047-1052.
- Peakall, R., Gilmore, S., Keys, W., Morgante, M. and Rafalski, A. (1998) Cross-species amplification of soybean (*Glycine max*) simple sequence repeats (SSRs) within the genus and other legume genera: implications for the transferability of SSRs in plants. *Mol. Biol. Evol.* **15**, 1275-1287.
- Pépin, L., Amigues, Y., Lépingle, A., Berthier, J.-L., Bensaid, A. and Vaiman, D. (1995) Sequence conservation of microsatellites between *Bos taurus* (cattle), *Capra hircus* (goat) and related species. Examples of use in parentage testing and phylogeny analysis. *Heredity* **74**, 53-61.
- Pfeiffer, A., Olivieri, A.M. and Morgante, M. (1997) Identification and characterization of microsatellites in Norway spruce (*Picea abies* K.). *Genome* **40**, 411-419.
- Phan, J., Reue, K. and Peterfy, M. (2000) MS-IRS PCR: a simple method for the isolation of microsatellites. *BioTechniques* **28**, 18-20.
- Powell, W., Morgante, M., McDevitt, R., Vendramin, G.G. and Rafalski, J.A. (1995a) Polymorphic simple sequence repeat regions in chloroplast genomes: Applications to the population genetics of pines. *Proc. Natl. Acad. Sci. USA* **92**, 7759-7763.
- Powell, W., Morgante, M., Andre, C., McNicol, J.W., Machray, G.C., Doyle, J.J., Tingey, S.V. and Rafalski, J.A. (1995b) Hypervariable microsatellites provide a general source of polymorphic DNA markers for the chloroplast genome. *Curr. Biol.* **5**, 1023-1029.
- Powell, W., Machray, G.C. and Provan, J. (1996) Polymorphism revealed by simple sequence repeats. *Trends Plant Sci.* **1**, 215-222.
- Prevost, A. and Wilkinson, M.J. (1999) A new system of comparing PCR primers applied to ISSR fingerprinting of potato cultivars. *Theor. Appl. Genet.* **98**, 107-112.

- Primmer, C.R., Möller, A.P. and Ellegren, H. (1996) A wide-range survey of cross-species microsatellite amplification in birds. *Mol. Ecol.* **5**, 365-378.
- Primmer, C.R. and Ellegren, H. (1998) Patterns of molecular evolution in avian microsatellites. *Mol. Biol. Evol.* **15**, 997-1008.
- Prochazka, M (1996) Microsatellite hybrid capture technique for simultaneous isolation of various STR markers. *Genome Res.* **6**, 646-649.
- Provan, J., Soranzo, N., Wilson, N.J., Goldstein, D.B. and Powell, W. (1999a) A low mutation rate for chloroplast microsatellites. *Genetics* **153**, 943-947.
- Provan, J., Thomas, W.T.B., Forster, B.P. and Powell, W. (1999b) Copia-SSR: a simple marker technique which can be used on total genomic DNA. *Genome* **42**, 363-366.
- Ramsay, L., Macaulay, M., Cardle, L., Morgante, M., Ivanissevich, S., Maestri, E., Powell, W. and Waugh, R. (1999) Intimate association of microsatellite repeats with retrotransposons and other dispersed repetitive elements in barley. *Plant J.* **17**, 415-425.
- Ramser, J., Weising, K., Chikaleke, V. and Kahl, G. (1997) Increased informativeness of RAPD analysis by detection of microsatellite motifs. *BioTechniques* **23**, 285-290.
- Ratnaparkhe, M.P., Santra, D.K., Tullu, A. and Muehlbauer, F.J. (1998) Inheritance of inter-simple-sequence-repeat polymorphisms and linkage with a fusarium wilt resistance gene in chickpea. *Theor. Appl. Genet.* **96**, 348-353.
- Richardson, T., Cato, S., Ramser, J., Kahl, G. and Weising, K. (1995) Hybridization of microsatellites to RAPD: a new source of polymorphic markers. *Nucleic Acids Res.* **23**, 3798-3799.
- Rico, C., Rico, I. and Hewitt, G. (1996) 470 million years of conservation of microsatellite loci among fish species. *Proc. Roy. Soc. Lond. B* **263**, 549-557.
- Riley, J., Butler, R., Ogilvie, D., Finier, R., Jenner, D., Powel, S., Anand, R., Smith, J.C. and Markham, A.F. (1990) *Nucleic Acids Res.* **18**, 2887-2890.
- Rithidech, K.N., Dunn, J.J. and Gordon, C.R. (1997) Combining multiplex and touchdown PCR to screen murine microsatellite polymorphisms. *BioTechniques* **23**, 36-44.
- Rodriguez, H., Geistlinger, J., Berlyn, G., Kahl, G. and Weising, K. (2000) Characterization of novel microsatellite loci isolated from the tropical dioecious tree *Simarouba amara*. *Mol. Ecol.* **9**, 489-504.
- Rohlf, F.J. (1993) NTSYS-pc: Numerical Taxonomy and Multivariate Analysis System. Exeter Software, Setauket, NY, USA
- Röder, M.S., Plaschke, J., König, S.U., Börner, A., Sorrells, M.E., Tanksley, S.D. and Ganal, M.W. (1995) Abundance, variability and chromosomal location of microsatellites in wheat. *Mol. Gen. Genet.* **246**, 327-333.
- Röder, M.S., Korzun, V., Wendehake, K., Plaschke, J., Tixer, M.H., Leroy, P. and Ganal, M.W. (1998) A microsatellite map of wheat. *Genetics* **149**, 2007-2023.

- Rosetto, M., McLauchlan, A., Harris, F.C.L., Henry, R.J., Baverstock, P.R., Lee, L.S., Maguire, T.L. and Edwards, K.J. (1999) Abundance and polymorphism of microsatellite markers in the tea tree (*Melaleuca alternifolia*, Myrtaceae). *Theor. Appl. Genet.* **98**, 1091-1098.
- Roy, M.S., Geffen, E., Smith, D., Ostrander, E.A. and Wayne, R.K. (1994) Patterns of differentiation and hybridization in North American wolflike canids, revealed by analysis of microsatellite loci. *Mol. Biol. Evol.* **11**, 553-570.
- Rubinsztein, D.C., Amos, W., Leggo, J., Goodburn, S., Jain, S., Li, S.H., Margolis, R.L., Ross, C.A. and Ferguson-Smith, M.A. (1995) Microsatellite evolution - evidence for directionality and variation in rate between species. *Nature Genet.* **10**, 337-343.
- Saghai-Marouf, M.A., Biyashev, R.M., Yang, G.P., Zhang, Q. and Allard, R.W. (1994) Extraordinarily polymorphic microsatellite DNA in barley: species diversity, chromosomal locations, and population dynamics. *Proc. Natl. Acad. Sci. USA* **91**, 5466-5470.
- Salimath, S.S., De Oliveira, A.C., Godwin, I.D. and Bennetzen, J.L. (1995) Assessment of genome origins and genetic diversity in the genus *Eleusine* with DNA markers. *Genome* **38**, 757-763.
- Sanchez de la Hoz, M.P., Dávila, J.A., Loarce, Y. and Ferrer, E. (1996) Simple sequence repeat primers used in polymerase chain reaction amplifications to study genetic diversity in barley. *Genome* **39**, 112-117.
- Santra, D.K., Tekeoglu, M., Rathnaparkhe, M.L. and Muehlbauer, F.J. (1998) Molecular markers for resistance to *Ascochyta* blight in chickpea. *Proceedings of the 3rd European Conference on Grain Legumes*, Valladolid, Spain pp 62-63.
- Schlötterer, C., Amos, B. and Tautz, D. (1991) Conservation of polymorphic simple sequence loci in cetacean species. *Nature* **354**, 63-65.
- Schlötterer, C., Ritter, R., Harr, B. and Brem, G. (1998) High mutation rate of a long microsatellite allele in *Drosophila melanogaster* provides evidence for allele-specific mutation rates. *Mol. Biol. Evol.* **15**, 1269-1274.
- Schmidt, T. and Heslop-Harrison, J.S. (1996) The physical and genomic organization of microsatellites in sugar beet. *Proc. Natl. Acad. Sci. USA* **95**, 8761-8765.
- Schug, M.D., Hutter, C.M., Wetterstrand, K.A., Gaudette, M.S., Mackay, T.F.C. and Aquadro, C.F. (1998) The mutation rates of di-, tri- and tetranucleotide repeats in *Drosophila melanogaster*. *Mol. Biol. Evol.* **15**, 1751-1760.
- Schwengel, D.A., Jedlicka, A.E., Nanthakumar, E.J. Weber, J.L. and Levitt, R.C. (1994). Comparison of fluorescence-based semi-automated genotyping of multiple microsatellite loci with autoradiographic techniques. *Genomics* **22**, 46-54.
- Scott, K.D., Eggler, P., Seaton, G., Rosetto, M., Ablett, E.M., Lee, L.S. and Henry, R.J. (2000) Analysis of SSRs derived from grape ESTs. *Theor. Appl. Genet.* **100**, 723-726
- Scott, L.J., Cross, M., Shepherd, M., Maguire, T. and Henry, R.J. (1999) Increasing the efficiency of microsatellite discovery from poorly enriched libraries in coniferous forest species. *Plant. Mol. Biol. Rep.* **17**, 351-354

- Scrimshaw, B.J. (1992) A simple nonradioactive procedure for visualization of (dC-dA)_n dinucleotide repeat length polymorphisms. *BioTechniques* **13**, 189
- Sharma, P.C., Winter P., Bünger, T., Hüttel, B., Weigand, F., Weising, K. and Kahl, G. (1995a) Abundance and polymorphism of di-, tri- and tetra-nucleotide tandem repeats in chickpea (*Cicer arietinum* L.). *Theor. Appl. Genet.* **90**, 90-96
- Sharma, P.C., Hüttel, B., Winter, P., Kahl, G., Gardner, R.C. and Weising, K. (1995b) The potential of microsatellites for hybridization- and polymerase chain reaction-based DNA fingerprinting of chickpea (*Cicer arietinum* L.) and related species. *Electrophoresis* **16**, 1755-1761
- Sia, E.A., Jink-Robertson, S. and Petes, T.D. (1997) Genetic control of microsatellite instability. *Mutation Res.* **383**, 61-70.
- Simon, C.J. and Muehlbauer, F.J. (1997) Construction of a chickpea linkage map and its comparison with maps of pea and lentil. *J. Hered.* **88**, 115-119.
- Smeets, H.J.M., Brunner, H.G., Ropers, H.-H. and Wieringa, B. (1989) Use of variable simple sequence motifs as genetic markers: application to study of myotonic dystrophy. *Hum. Genet.* **83**, 245-251.
- Smulders, M.J.M., Bredemeijer, G., Rus-Kortekaas, W., Arens, P. and Vosman, B. (1997) Use of short microsatellites from database sequences to generate polymorphisms among *Lycopersicon esculentum* cultivars and accessions of other *Lycopersicon* species. *Theor. Appl. Genet.* **94**, 264-273.
- Sonnante, G., Stockton, T., Nodari, R.O., Becerra-Velasquez, V.L. and Gepts, P. (1994) Evolution of genetic diversity during the domestication of common bean (*Phaseolus vulgaris* L.). *Theor. Appl. Genet.* **89**, 629-635.
- Soranzo, N., Provan, J. and Powell, W. (1999) An example of microsatellite length variation in the mitochondrial genome of conifers. *Genome* **42**, 158-161.
- Stephan, W. and Kim, Y. (1998) Persistence of microsatellite arrays in finite populations. *Mol. Biol. Evol.* **15**, 1332-1336.
- Stephenson, P., Bryan, G., Kirby, J., Collins, A., Devos, K., Busso, C. and Gale, M. (1998) Fifty new microsatellite loci for the wheat genetic map. *Theor. Appl. Genet.* **97**, 946-949.
- Tachida, H. and Iizuka, M. (1992) Persistence of repeated sequences that evolve by replication slippage. *Genetics* **131**, 471-478.
- Tanksley, S.D., Ganai, M.W. and Martin, G.B. (1995) Chromosome landing: a paradigm for map-based gene cloning in plants with large genomes. *Trends Genet.* **11**, 63-68.
- Taramino, G. and Tingey, S. (1996) Simple sequence repeats for germplasm analysis and mapping in maize. *Genome* **39**, 277-287.
- Tautz, D. (1989) Hypervariability of simple sequences as a general source for polymorphic DNA markers. *Nucleic Acids Res.* **17**, 6463-6471.

- Temnykh, S., Park, W.D., Ayres, N., Cartinhour, S., Hauck, N., Lipovich, L., Cho, Y.G., Ishii, T. and McCouch, S.R. (2000) Mapping and genome organisation of microsatellite sequences in rice. *Theor. Appl. Genet* **100**, 697-712.
- Terauchi, R. and Konuma, A. (1994) Microsatellite polymorphism in *Dioscorea tokoro*, a wild yam species. *Genome* **37**, 794-801.
- Thomas, M.R., Cain, P. and Scott, N.S. (1994) DNA typing of grapevines: a universal methodology and database for describing cultivars and evaluating genetic relatedness. *Plant Mol. Biol.* **25**, 939-949.
- Thompson, L.E., Kepinski, S., Hanappel, U., Rae, S.J., Moule, C.L., Owen, P. and Edwards, K.J. (1999) An improved strategy for the development of high numbers of microsatellite markers from a wide range of plant species. In: *International Plant and Animal Genome Conference VII*. San Diego, California, p 447.
- Udupa, S.M., Sharma, A., Sharma, R.P. and Pai, R.A. (1993) Narrow genetic variability in *Cicer arietinum* L. as revealed by RFLP analysis. *J. Plant Biochem. Biotechnol.* **2**, 83-86.
- Ueno, S., Yoshimaru, H., Tomaru, N. and Yamamoto, S. (1999) Development and characterization of microsatellite markers in *Camellia japonica* L. *Mol. Ecol.* **8**, 335-336.
- van de Wiel, C., Arens, P. and Vosman, B. (1999) Microsatellite retrieval in lettuce (*Lactuca sativa* L.). *Genome* **42**, 139-149.
- van der Maesen, L.J.G. (1987) Origin, history and taxonomy of chickpea. In: Saxena MJ, Singh KB (eds): *The Chickpea*. Cambridge CAB International pp 11-34.
- Van Treuren, R., Kuittinen, H., Kärkäinen, K., Baena-Gonzalez, E. and Savolainen, O. (1997) Evolution of microsatellites in *Arabis petraea* and *Arabis lyrata*, outcrossing relatives of *Arabidopsis thaliana*. *Mol. Biol. Evol.* **14**, 220-229.
- Vendramin, G.G., Anzidei, M., Madaghiele, A. and Bucci, G. (1998) Distribution of genetic diversity in *Pinus pinaster* Ait. as revealed by chloroplast microsatellites. *Theor. Appl. Genet.* **97**, 456-463.
- Vogel, J.M. and Scolnik, P.A. (1997) Direct amplification from microsatellites: detection of of simple-sequence repeat-based polymorphism without cloning. In: Caetano-Anollés G, Gresshoff PM (eds): *DNA markers: Protocols, Applications and Overviews*. Wiley-Lis, John Wiley and Sons Inc, New York, pp 133-150.
- Vos, P., Hogers, R., Bleeker, M., Reijans, M., van de Lee, T., Hoernes, M., Frijters, A., Pot, J., Peleman, J., Kuiper, M. and Zabeau, M. (1995) AFLP: A new technique for DNA fingerprinting. *Nucleic Acids Res.* **23**, 4407-4414.
- Vosman, B. and Arens, P. (1997) Molecular characterization of GATA/GACA microsatellite repeats in tomato. *Genome* **40**, 25-33.
- Waldbieser, G.C. (1995) PCR-based identification of AT-rich tri- and tetranucleotide repeat loci in an enriched plasmid library. *BioTechniques* **19**, 742-744.

- Wallner, E., Weising, K., Rompf, R., Kahl, G. and Kopp, B. (1996) Oligonucleotide fingerprinting and RAPD analysis of *Achillea* species: characterization and long-term monitoring of micropropagated clones. *Plant Cell Reports* **15**, 647-652.
- Wang, Z., Weber, J.L., Zhong, G. and Tanksley, S.D. (1994) Survey of plant short tandem DNA repeats. *Theor. Appl. Genet.* **88**, 1-6.
- Wang, X. and Bowen, B. (1998) A progress report on corn genome projects at pioneer hybrid. In: *Plant and Animal Genome Conference VI*, San Diego, California
- Wang, D.G. et al. (1998) Large-scale identification, mapping, and genotyping of single-nucleotide polymorphisms in the human genome. *Science* **280**, 1077-1082.
- Weber, J.L. (1990) Informativeness of human (dC-dA)_n x (dG-dT)_n polymorphisms. *Genomics* **7**, 524-530.
- Weber, J.L. and May, P.E (1989) Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. *Am. J. Hum. Genet.* **44**, 388-396
- Weber, J.L. and Wong, C. (1993) Mutation of short tandem repeats. *Hum. Mol. Genet.* **2**, 1123-1128.
- Weising, K. and Gardner, R.C. (1999) A set of conserved PCR primers for the analysis of simple sequence repeat polymorphisms in chloroplast genomes of dicotyledonous angiosperms. *Genome* **42**, 9-19.
- Weising, K. and Kahl, G. (1997) Hybridization-based microsatellite fingerprinting of plants and fungi. In: Caetano-Anollés G, Gresshoff P (eds.) *DNA Markers: Protocols, Applications and Overviews*. Wiley-Liss, New York, pp27-54
- Weising, K., Weigand, F., Driesel, A.J., Kahl, G., Zischler, H. and Epplen, J.T. (1989) Polymorphic simple GATA/GACA repeats in plant genomes. *Nucleic Acids Res.* **17**, 10128.
- Weising, K., Ramser, J., Kaemmer, D., Kahlm G. and Epplen, J.T. (1991) Oligonucleotide fingerprinting in plants and fungi. In: Burke T, Dolf G, Jeffreys AJ, Wolff R (eds) *DNA fingerprinting: Approaches and applications*. Birkhäuser, Basel, pp 313-329.
- Weising, K., Kaemmer, D., Weigand, F., Epplen, J.T. and Kahl, G. (1992) Oligonucleotide fingerprinting reveals probe-dependent levels of informativeness in chickpea (*Cicer arietinum* L.). *Genome* **35**, 436-442.
- Weising, K., Nybom, H., Wolff, K. and Meyer, W. (1995a) DNA fingerprinting in plants and fungi. CRC Press, Boca Raton, Florida
- Weising, K., Atkinson, R.G. and Gardner, R.C. (1995b) Genomic fingerprinting by microsatellite-primed PCR: a critical evaluation. *PCR Methods & Applications* **4**, 249-255.
- Weising, K., Fungm R.W.M., Keeling, J., Atkinson, R.G. and Gardner, R.C. (1996) Cloning and characterization of microsatellite repeats from *Actinidia chinensis*. *Mol. Breeding* **2**, 117-131.
- Weising, K., Winter, P., Hüttel, B. and Kahl, G. (1998) Microsatellite markers for molecular breeding. *J. Crop Prod.* **1**, 113-143.

- Weissenbach, J., Gyapay, G., Dib, C., Vignal, A., Morissette, J., Millasseau, P., Vaysseix, G., Lathrop, M. (1992) A second-generation linkage map of the human genome. *Nature* **359**, 794-801.
- Westman, A.L. and Kresovich, S. (1998) The potential for cross-taxa simple-sequence repeat (SSR) Amplification between *Arabidopsis thaliana* L. and crop brassicas. *Theor. Appl. Genet.* **96**, 272-281.
- White, G. and Powell, W. (1997a) Isolation and characterization of microsatellite loci in *Swietenia humilis* (Meliaceae), and endangered tropical hardwood species. *Mol. Ecol.* **6**, 851-860.
- White, G. and Powell, W. (1997b) Cross-species amplification of SSR loci in the Meliaceae family. *Mol. Ecol.* **6**, 1195-1197.
- Whitton, J., Rieseberg, L.H. and Ungerer, M.C. (1997) Microsatellite loci are not conserved across the *Asteraceae*. *Mol. Biol. Evol.* **14**, 204-209.
- Wierdl, M., Dominska, M. and Petes, T.D. (1997) Microsatellite instability in yeast: dependance on the length of the microsatellite. *Genetics* **146**, 769-779.
- Williams, J.G.K., Kubelik, A.R., Livak, K.J., Rafalski, J.A. and Tingey, S.V. (1990) DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. *Nucleic Acids Res.* **18**, 6531-6535.
- Winter, P. and Kahl, G. (1995) Molecular marker technologies for plant improvement. *World J. Microbiol. Biotechnol.* **11**, 438-448.
- Winter, P., Pfaff, T., Udupa, S.M., Hüttel, B., Sharma, P.C., Sahi, S., Arreguin-Espinoza, R., Weigand, F., Muehlbauer, F.J. and Kahl, G. (1999) Characterization and mapping of sequence-tagged microsatellite sites in the chickpea (*Cicer arietinum* L.) genome. *Mol. Gen. Genet.* **262**, 90-101.
- Winter, P., Benko-Iseppon, A.-M., Hüttel, B., Ratnaparkhe, M., Tullu, A., Sonnante, G., Pfaff, T., Tekeoglu, M., Santra, D., Sant, V.J., Rajesh, P.N., Kahl, G. and Muehlbauer, F.J. (2000) A linkage map of the chickpea (*Cicer arietinum* L.) genome based on recombinant inbred lines from a *C. arietinum* x *C. reticulatum* cross: Localization of resistance genes for *Fusarium* races 4 and 5. *Theor. Appl. Genet.* **101**, 1155-1163.
- Winzeler, E.A. et al. (1998) Direct allelic variation scanning of the yeast genome. *Science* **281**, 1194-1197.
- Wolfe, A.D., Xiang, Q.Y. and Kephart, S.R. (1998) Assessing hybridization in natural populations of *Penstemon* (Scrophulariaceae) using hypervariable intersimple sequence repeat (ISSR) bands. *Mol. Ecol.* **7**, 1107-1125.
- Wolff, K. and Morgan-Richards, M. (1998) PCR markers distinguish *Plantago major* subspecies. *Theor. Appl. Genet.* **96**, 282-286.
- Wolff, K., Zietkiewicz, E. and Hofstra, H. (1995) Identification of chrysanthemum cultivars and stability of fingerprint patterns. *Theor. Appl. Genet.* **91**, 439-447.

Wu, K.-S. and Tanksley, S.D. (1993) Abundance, polymorphism and genetic mapping of microsatellites in rice. *Mol. Gen. Genet.* **241**, 225-235.

Wu, K., Jones, R., Danneberger, L. and Scolnik, P.A. (1994) Detection of microsatellite polymorphisms without cloning. *Nucleic Acids Res.* **22**, 3257-3258.

Yamamoto, K. and Sasaki, T. (1997) Large-scale EST sequencing in rice. *Plant Mol. Biol.* **35**, 135-144.

Zhao, X. and Kochert, G. (1993) Phylogenetic distribution and genetic mapping of a (GCC)_n microsatellite from rice (*Oryza sativa* L.). *Plant Mol. Biol.* **21**, 607-614.

Ziegle, J.S., Su, Y., Corcoran, K.P., Nie, L., Mayrand, P.E., Hoff, L.B., McBride, L.J., Kronick, M.N. and Diehl, S.R. (1992) Application of automated DNA sizing technology for genotyping microsatellite loci. *Genomics* **14**, 1026-1031.

Zietkiewicz, E., Rafalski, A. and Labuda, D. (1994) Genome fingerprinting by simple sequence repeat (SSR)-anchored polymerase chain reaction amplification. *Genomics* **20**, 176-183.

Zischler, H., Kammerbauer, C., Studer, R., Grzeschik, K.H. and Epplen, J.T. (1992) Dissecting (CAC)₅/GTG₅ multilocus fingerprints from man into individual locus-specific, hypervariable components. *Genomics* **13**, 983-990.

Zohary, D. and Hopf, M. (1993) Domestication of plants in the old world. Clarendon Press, Oxford. pp 1-271.

DEVELOPMENT AND APPLICATION OF MOLECULAR MARKERS IN CONIFERS

R. SCHUBERT AND G. MÜLLER-STARCK

*Technical University of Munich, Center of Life and Food Sciences Weihenstephan, Department of Plant Sciences, Section of Forest Genetics
Am Hochanger 13, D-85354 Freising, Germany*

1. INTRODUCTION

1.1. Forests of the World

Natural forests and forest plantations currently cover 26.6% of the total land area of the earth (FAO 1997). Forests basically affect both, carbon and water cycles on our planet and provide manifold benefits by harboring species and protecting landscapes against erosion, by supplying shelter against wind, snow, radiation or noise, and by offering various recreation purposes and outdoor activities. Economic benefits arise from utilization of timber, bark and many other natural products for various industrial purposes (mainly pulp and paper production), for energy production and for supply of food.

Despite the fact that some tree species were cultivated for several thousand years, 97% of the world's forests still represent natural or semi-natural forests (FAO 1999), which are considered to be largely undomesticated in comparison with annual and perennial species: Due to the long generation cycles of tree species, their gene pools have not been intensively managed by breeding technologies and therefore still contain a naturally existing level of biodiversity. Since long-lived tree populations moreover exist in a wide range of habitats, they are exposed to complex and unpredictable biotic and abiotic stress components. Control of stress is, however, not possible or strictly limited in forests in contrast to crop plant systems. Tree populations are vulnerable particularly because of the discrepancy between extremely short generation cycles of mobile parasites in contrast to long generation cycles of immobile hosts.

Survival of forest tree populations strongly depends on its ability to adapt to heterogeneous environmental conditions. Compared to populations of annual and perennial plant species, forest tree populations reveal high levels of genetic variation (e.g. Mitton 1983; Hamrick and Godt 1989; Müller-Starck *et al.* 1992), which have been verified to be correlated with fitness (e.g. Mitton and Grant 1984; Allendorf and Leary 1986) and are considered to predominantly determine the adaptive abilities of populations (e.g. Gregorius 1991).

Generally, the application of breeding technologies is hampered by the longevity of tree species and the necessity of maintaining high levels of genetic variability in both, breeding and productive populations.

1.2 Two Taxonomic Groups of Forest Tree Species

The world's forests are almost equally subdivided in boreal/temperate forests and tropical/subtropical forests dominated by conifers and broad-leaf species, respectively. The large taxonomic group of broad-leaf species represents angiospermous seed plants. These species have been radiating in the last 120 million years. On the other hand, the small group of coniferous tree species (gymnosperms) has arisen during 300 million years of evolution (for detailed history of tracheophyte evolution see Niklas *et al.* 1983). The amplification and dispersal of genes to form complex families (Kinlaw and Neale 1997), including intragenic small scale duplications (Perry and Furnier 1996; Schubert *et al.* 1998) and the evolution of archaic pseudogene clades (Kvarnheden *et al.* 1998), appears to be especially prominent in conifers and could be responsible for the extraordinary large genome sizes associated with gymnosperms (approx. 30 000 Megabases per nucleus according to Govindaraju and Cullis 1991). Interestingly, conifers confirm retardations in evolutionary rates when single nuclear genes have been aligned with homologous sequences from angiosperms (Yokoyama and Harry 1993; MacKay *et al.* 1995; Häger and Fischer 1999). Based on distance matrix analysis and parsimony analysis, nucleotide alignments have furthermore shown that genes of gymnosperms gave rise to monophyletic clades distinct from those of the homologous clades of angiospermous genes, reflecting the phylogenetic gap between both groups of seed plants (Schubert *et al.* 1997; Schubert *et al.* 1998).

2. MOLECULAR MARKERS IN USE OF CONIFERS

Until the 1990s, various markers have been employed for population genetic surveys, indicating genetic diversity exclusively at the level of ex-

pressed phenotypic traits. Such markers are morphological characters, isoenzymes, immunoproteins, terpenes and fatty acids (e.g. Lagercrantz and Ryman 1990; Müller-Starck 1991; Prus-Glowacki 1982; Baradat *et al.* 1991; Wolff *et al.* 1997). Despite the large amount of existing genetic data, particularly based on isoenzyme gene markers, the main limitation of these classical markers is the fact that they are restricted in number. After the optimization of DNA extraction protocols (e.g. Ziegenhagen *et al.* 1993), an ongoing characterization of coniferous nuclear and extra-nuclear genome components was mainly achieved by PCR-based methods, focusing on some widely used economically and ecologically important pine and spruce species. Resulting from this pioneering work, different types of molecular markers became available for monitoring genetic variability directly at the nucleotide level in such genomic regions which were not previously characterized by the classical marker systems above mentioned.

Development and application of modern molecular marker techniques have greatly facilitated research on genome organization, phylogeny reconstruction, species speciation, clone identification, analysis of biodiversity and mating systems. Nevertheless, current knowledge with respect to coniferous genomes is still fragmentary and lags behind corresponding investigations of angiospermous genomes. Taken into account the complete genome sequence recently reported by “The *Arabidopsis* Genome Initiative” (2000), there is no comparable sequencing project established so far to attempt a whole genome characterization of a coniferous tree species. Current genome maps of conifers coalesce into more than the expected number of linkage groups, demonstrating the experimental difficulties encountered when attempting to saturate a large and complex genome with molecular markers (e.g. Paglia *et al.* 1998).

2.1 Nuclear DNA Markers: Codominant vs. Dominant Inheritance

Restriction fragment length polymorphism (**RFLP**) represents the first DNA based marker technology which has been successfully established for conifers (Devey *et al.* 1991; Bobola *et al.* 1992). Since this technique requires laborious methods of marker detection, including Southern blot hybridization, RFLPs are now commonly replaced by more rapid applications of **PCR** (polymerase chain reaction). This technology starts with low quantities of tree DNA and provides a visible marker fragment under the direction of a thermostable DNA polymerase and oligonucleotide primers. With respect to a high-throughput screening of variation, primers should be selected for their ability to generate polymorphic markers which are scorable directly by using both standard agarose gel and polyacrylamide gel electro-

phoresis without additional manipulations. To identify sequence variations at the level of single base point mutations between nucleic acid samples, however, post-PCR manipulations have been conducted, including restriction cleavage and sequence-dependent gel mobility shifts (e.g. Burban *et al.* 1999).

Microsatellites (simple sequence repeats) are stretches of genomic DNA, consisting of perfect, imperfect as well as compound tandem repeats of mono-, di-, tri-, tetra- or penta-nucleotides (for detailed information see Powell *et al.* 1996), which are predicted to harbor high levels of polymorphism due to slippage-mispairing occurring in the process of DNA replication and unequal crossing-over. Based on random genomic libraries and cDNA libraries, microsatellite units were identified among the recombinant clones by hybridization with oligomeric repeat-containing probes. Most of the discovered microsatellites are localized within anonymous noncoding genomic regions, but trinucleotide repeats can be associated with the reading frame of protein encoding gene parts. Primer pairs, complementing the flanking unique microsatellite regions were designed after sequence analysis and the marker fragments obtained from genomic PCR have been assessed because of their variability (Echt *et al.* 1996; van de Ven and McNicol 1996; Pfeiffer *et al.* 1997; Scotti *et al.* 2000a; Elsik *et al.* 2000; Lian *et al.* 2000).

Until the early 1990s, the approach to generate locus-specific primer pairs was hindered in case of coniferous tree species by the absence of gene sequences in public databases. Consequently, particular cDNA sequencing projects were initiated and primer sequences have been retrieved from the dataset to amplify sequence-tagged-sites (also termed expressed-sequence-tags). The resultant multiallelic **STS/EST** markers (Perry and Bousquet 1998; Schubert *et al.* 2001; Temesgen *et al.* 2001) as well as the mentioned microsatellite markers were utilized in a variety of applications, mainly focusing on genome mapping and population studies.

Nowadays, the variability of the long-terminal repeat sequences of *Ty1*-copia group **retrotransposons** seems to become a source of powerful molecular markers, even in conifers (Pearce *et al.* 1999). Since such mobile genetic elements are known to being distributed throughout the whole genome of the host plant (for detailed informations see Brandes *et al.* 1997), their pattern of variation makes them ideal for integrating genetic maps derived from related crosses and to establish markers for such genomic regions which are currently not captured by markers.

The nuclear DNA markers mentioned above as well as isoenzyme gene markers are inherited in codominant fashion, which ensures for diploid conifers the identification of both alleles occurring at the Mendelian locus tested. Once developed, codominant markers can be considered the most reliable tool to assess the actual level of homozygosity, a basic parameter in population genetics (reviewed by Sunnucks 2000). Performing the exact test

of Hardy-Weinberg proportion for the multiple alleles observed at a single genetic locus (Guo and Thompson 1992), excess of homozygotes or heterozygotes indicates that the analyzed population has no random-mating structure and the locus under investigation could therefore indicate for inbreeding and/or various forms of selection.

Single gene probes and target-specific primer pairs have been shared between related gymnosperms (Ahuja *et al.* 1994; Tsumura *et al.* 1997; Perry *et al.* 1999), reflecting the slow-down of gene evolutionary rates mentioned. Nevertheless, the development of additional codominant DNA markers for tagging the complex genomes of conifers still represents a time-consuming approach with high operational costs. Marker technologies based on arbitrarily chosen PCR primer sequences therefore indicate a more efficient alternate: In terms of random amplified polymorphic DNA (**RAPD**) analysis (Williams *et al.* 1990), anonymous multilocus marker bands are synthesized after the annealing of an universal 10 base pair primer with the genomic target. The efficiency of this marker technology has been successfully tested in a variety of applications for coniferous tree species, including rapid chromosomal mapping based on haploid megagametophytes (Tulsiaram *et al.* 1992; Binelli and Bucci 1994), evaluating the genetic stability of somatic embryogenesis-derived populations (Isabel *et al.* 1993) and fingerprinting of clones and provenances (Scheepers *et al.* 1997).

The other available multilocus profiling technique which starts with primers of arbitrary nucleotide sequence, known as amplified fragment length polymorphic DNA (**AFLP**), has been originally developed by Vos and co-workers (1995) and has been modified for successfully analyzing angiosperm and gymnosperm tree species (Cervera *et al.* 2000). Based on a combination of different restriction endonucleases and primer sequences, this technology yields a high number of scorable bands to attempt a whole-genome characterization via markers. Investigating embryonic stage inbreeding depression in loblolly pine, an essentially complete genome map has been constructed using AFLP markers (Remington and O'Malley 2000).

Each RAPD and AFLP marker band is scorable by its presence or absence (Carlson *et al.* 1991), indicating a homozygous individual when the marker band is lacking. Taken into account this dominant inheritance, the fundamental limitation of RAPD and AFLP technologies is the fact that homozygous trees and heterozygous individuals cannot be distinguished in the presence of the marker band which strongly interferes with genotypic arrays. Since RAPD markers provide decreased reliability and transferability across laboratories due to the reduced annealing conditions required, single RAPD markers have been converted to locus-specific dominant **SCAR** (sequence-characterized amplified region) markers following sequence analysis and primer design (Scotti *et al.* 1998).

2.2 Markers for Organelle Type DNA

In contrast to the biparental inheritance of nuclear genes, chloroplast and mitochondrion organelles of conifers are inherited paternally and maternally, respectively, as has been discovered by analyzing polymorphic RFLP markers in the progenies of controlled crosses (e.g. Wagner *et al.* 1989; Stine and Keathley 1990; Sutton *et al.* 1991; Marshall and Neale 1992). Using tree samples from introgression zones, RFLP markers have been employed to detect hybridization between related coniferous tree species (Sutton *et al.* 1991; Bobola *et al.* 1996). Mitochondrial RFLPs were also successfully tested in population studies (Strauss *et al.* 1993).

With respect to the chloroplast genome of conifers, the development of modern marker technologies has substantially benefited from the chloroplast sequence which has been previously reported for *Pinus thunbergii* (Wakasugi *et al.* 1994). This full-length sequence allowed to identify repeat units and to design primer sequences complementing the flanking regions of the discovered microsatellites (Vendramin *et al.* 1996). The retrieved universal primer sequences ensure the analysis of homologous polymorphic microsatellite regions in a number of species (Vendramin and Ziegenhagen 1997; Echt *et al.* 1998; Vendramin *et al.* 2000), indicating the reduced evolutionary rate of the chloroplast genome among different species of the *Pinaceae*. A slow-down in the evolutionary rate of the chloroplast genome has been suggested by an universal primer pair, tracing a single restriction site polymorphism for the *psbC* gene in a wide range of 62 woody gymnosperms and dicotyledonous angiosperms (Ziegenhagen and Fladung 1997). Polymorphic tandem units of mitochondrial DNA have been rarely used in population studies of conifers until now (see topic 3.2.).

The selection of appropriate markers is essential since no marker currently used in forest tree species can be expected to be universal in answering biological questions. On the other hand, the increasing number and variety of genetic markers makes it difficult for end-users of this technology to select suitable genetic markers. Guidelines for choosing markers, including sampling strategies and available software for marker analysis, were compiled within the EU-project BIO 4-CT96-0706 “*Development, optimization and validation of molecular tools for assessment of biodiversity in forest trees*” (<http://webdoc.sub.gwdg.de/ebook/y/1999/whichmarker/index.htm>).

3. CASE STUDIES

This chapter covers case studies in order to illustrate important principles of the development and application of DNA marker technologies in conifers, including their impact on population genetics and tree breeding.

3.1 Marker Development Based on Genomic Sequences Encoding Cinnamyl Alcohol Dehydrogenase

Cinnamyl alcohol dehydrogenase (CAD) catalyses the last step in the biosynthesis of the monomeric precursors of lignin (for detailed informations see Boudet *et al.* 1995; Grima-Pettenati and Goffner 1999). We have previously isolated three highly uniform genomic CAD sequences from the existing multigene family in Norway spruce [*Picea abies* (L.) Karst.] and have furthermore detected a polymorphic region among the aligned sequences of the very C-terminal localized intron no. V, which is characterized by two imperfect tandem repeats of 29 bp and 49 bp (Schubert *et al.* 1998).

To obtain a CAD gene marker, the PCR primer pair P3 (5'-GAAAAAATCTGCCTTCAGTAAACC3') / P4 (5'-CGATGAAACTCCCAGCTATGC3') was synthesized (Schubert *et al.* 1998), complementing the flanking regions of the mentioned repeat motifs. On re-examination of primer pair P3/P4 using a modified PCR program and Platinum[®] *Taq* DNA polymerase (provided by LifeTechnologies, Karlsruhe, Germany), an enzyme becoming hot-start activated during the initial denaturation step of the PCR cyclus (5 minutes at 94°C), Riegel (2001) has found two different amplicons in the proved diploid bud sample (marked by symbols I and O, see Figure 1). Both marker bands revealed a 1:1 segregation when the corresponding individual megagametophyte samples (endosperm tissues) of the same test tree were analyzed by PCR as shown in Figure 1. Haploid megagametophytes represent the female meiotic products in conifers, indicating that the PCR product patterns obtained under the direction of primer pair P3/P4 are in accordance with Mendelian mode of inheritance. For the heterozygous tree analyzed, the polymorphic marker is therefore distinguishing between both different alleles of the same CAD gene locus.

This result is strongly contrasted by the originally published P3/P4-derived PCR product patterns which had clearly demonstrated the existence of two different CAD gene loci (Schubert *et al.* 1998). The more stringent PCR conditions used now are obviously able to discriminate between the highly uniform CAD gene loci of Norway spruce, ensuring the amplification of a single locus. We currently identified 17 alleles for this new codominant

marker by analyzing a test population of 110 trees (Forest of Kranzberg, Bavaria), showing that the PCR-amplified repeat motif region of intron no. V harbors high genetic variability. The codominant inheritance of the CAD gene marker was further confirmed by analyzing the progeny of a controlled cross (Riegel 2001).

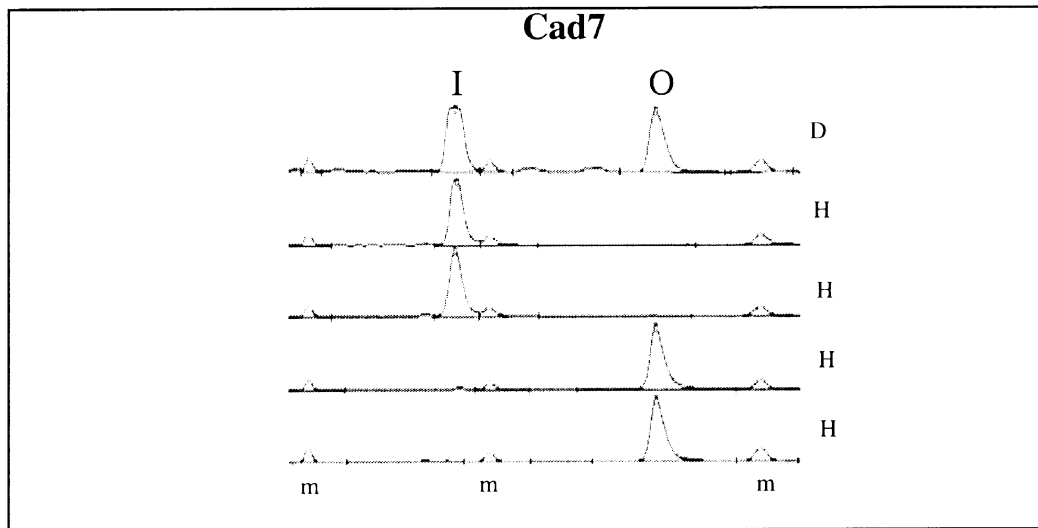


Figure 1. Segregation of alleles I and O at a codominant CAD gene locus of Norway spruce, among haploid (H) megagametophytes of a heterozygous tree (diploid bud sample indicated by symbol D). Internal size standards (m) were added after PCR (data taken from Riegel 2001). For PCR conditions (program JAN1), electrophoresis and identification of fluorescently-labelled products by the ALFexpress instrument see Schubert *et al.* 2001.

3.2 Monitoring Interpopulational and Intrapopulational Genetic Variation

The assessment of genetic variability within and among tree populations is essential in population genetics in order to obtain information on the history of coniferous species, the characterization of the present gene pools, the evaluation of human impacts on the natural level of genetic diversity, and the utilization with respect to tree breeding. Multiallelic markers, confirming the high genetic variation usually associated with tree species, are therefore most desirable tools.

Based on isoenzyme gene markers, 90-99% of the measured genetic diversity exists within populations, while the remaining small amount of genetic diversity covers interpopulational variation (for survey see Müller-Starck and Ziehe 1991). The interpopulational variation detected by RAPD markers (DeVerno and Mosseler 1997) and SCAR markers (Scotti *et al.* 2000b) is comparable with isoenzyme data or reveals not much higher values, respectively. When affected by diversifying selection or geographic

isolation, surveys of nuclear marker may provide, however, important informations about the fragmentation and expansion of conifers (Latta and Mitton 1997; Scotti *et al.* 2000b). The small interpopulational differentiation reported from nuclear markers reflects that outbreeding coniferous tree species intensively interact within their large continuous habitats via gene flow, resulting in homogenizing effects of the genetic information. With respect to the reproduction of natural stands and seed orchards, hypervariable microsatellite markers of conifers associated with extraordinary high levels of genetic variability are expected to become powerful indicators for monitoring the actual level of pollen migration and to identify maternal and paternal parents of an offspring, as has been previously documented for angiospermous tree species (White and Powell 1997; Lexer *et al.* 2000). For Norway spruce, Scotti and co-workers (2000a) detected pollen contamination in the F_1 progeny of a controlled cross by the identification of nuclear microsatellite marker alleles which are completely absent in the germplasm of both parents. The potential usefulness of chloroplast microsatellites as markers for paternity analysis has been demonstrated in silver fir (Ziegenhagen *et al.* 1998).

In contrast to nuclear markers, the genetic diversity verified by organellar haplotypes can be expected to differentiate more strongly between tree populations, rather than indicating intrapopulational variation, due to the different evolutionary rates and inheritance types of nuclear loci and organelle-localized genes. As a consequence of the highly informative interpopulational differentiation associated with chloroplast markers (e.g. Echt *et al.* 1998; Bucci and Vendramin 2000), different phylogeographic gene pools and re-colonisation pathways were identified for conifers, according to the existence of distinct refugia, also known from pollen records, where the respective tree species survived the last glacial period. The identification of such genetically homogenous regions has important economic relevance for the forest management in order to avoid the cultivation of inappropriate foreign provenances which are not adapted at the local site conditions and to develop meaningful conservation strategies for species currently endangered by genetic erosion. Based on estimates of seed and pollen flow which have been derived from mitochondrial DNA and chloroplast DNA markers, respectively, Latta and Mitton (1997) found a much higher differentiation for maternally inherited mitochondrial DNA polymorphisms than for paternally chloroplast DNA in limber pine. This result demonstrates that efficient gene flow by wind-dispersed pollen compensates genetic differentiation among continuous populations of conifers. For Norway spruce, two polymorphic tandem repeats have been recently discovered in the second intron of the mitochondrial *nad1* gene (Sperisen *et al.* 2001). The most common size variants detected for this marker in a survey of 940 trees, representing 90 European populations, reveal pronounced population differentiation and

a distinct geographical distribution based on the phylogenetic separation of two putative glacial refugia, which are localized in the Moscow area and the Carpathians or the Balkan Peninsula, respectively (see Figure 2).

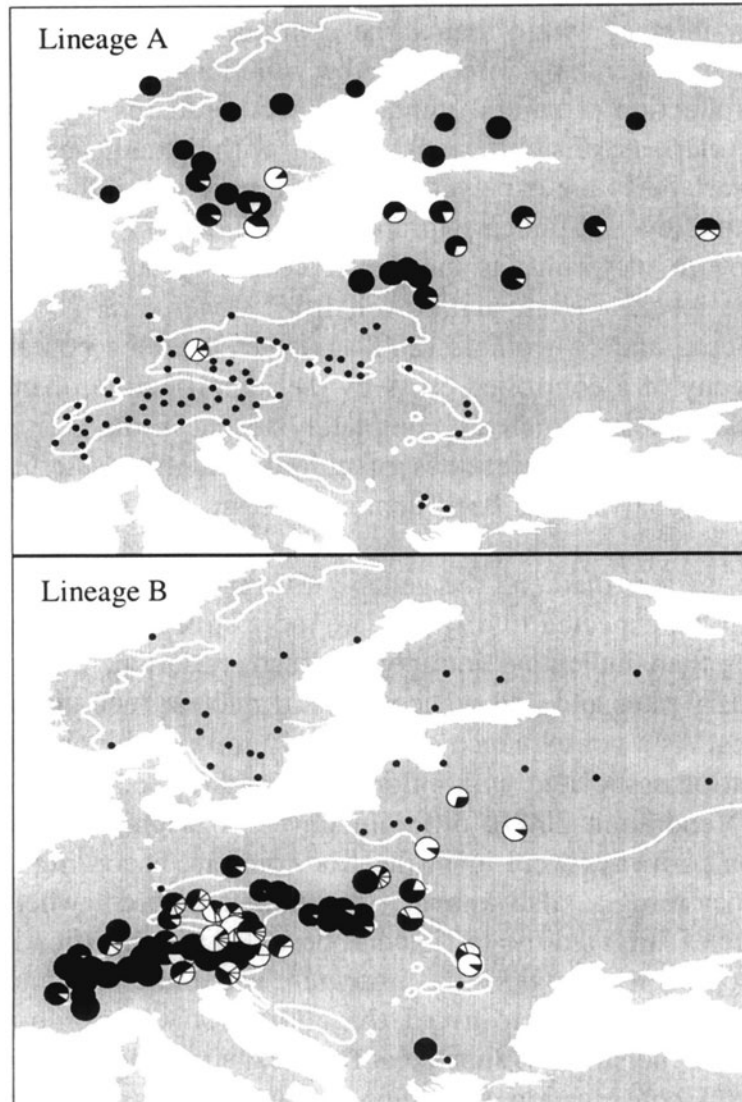


Figure 2. Geographical distribution of most abundant mitochondrial size variants, representing lineages A and B, of *Picea abies* within its natural European range. The relative frequency of each size variant is given by the black sections of circles. Dots indicate positions of sampled populations not containing the respective size variant (graphs taken from Sperisen *et al.* 2001).

For codominant markers, various types of polymorphisms were found by analyzing the frequencies of present alleles. Most loci indicate minor polymorphisms by the predominant frequency of one allele in combination with additional alleles which appear in very low frequencies. Low-frequency alleles are considered to be essential parts of a genetic resource, ensuring the

survival of natural populations under the conditions of a dynamic environment (e.g. Finkeldey and Gregorius 1991). For fragmented species endangered by genetic erosion, appropriate conservation programs therefore try to protect the owners of seldom alleles (for principles and practice in forest conservation genetics see Young *et al.* 2000).

TABLE 1. Summary of six newly developed codominant EST markers of Norway spruce (for gene identification, primer sequences, PCR conditions, and allele identification see Schubert *et al.* 2001)

Marker (Gene locus)	Putative Identification	Detection of polymor- phisms	Frequencies of present alleles
PA0002	A-like cyclin	PCR product polymorphic in length after HaeIII cut	A: 50%; B: 50%
PA0034	Nonidentified gene with imperfect repeat motif	PCR product polymorphic in length	A: 5%; B: 7%; C: 71%; D: 5%; E: 12%
PA0038	Halotolerance protein	PCR product polymorphic in length after HinfI cut	A: 4%; B: 53%; C: 43%
PA0043	High-molecular heat-shock protein	PCR product polymorphic in length	A: 55%; B: 16%; C: 24%; D: 5%
PA0055	Beta-subunit of the ATP synthase complex	PCR product polymorphic in length after DraI cut	A: 72%; B: 28%
PA0066	60S-ribosomal protein L13-2	PCR product polymorphic in length	A: 2%; B: 74%; C: 1%; D: 23%

In contrast to isoenzyme gene markers, we found a trend to more even frequency distributions when six newly developed codominant EST markers were tested using a natural population of 110 trees (Schubert *et al.* 2001). The tendency towards minor polymorphisms was only exhibited by three molecular markers (PA0034, PA0055 and PA0066; see TABLE 1) while marker PA0002 shows a typical major polymorphisms, i.e. an even distribution of both present alleles. Moreover, markers PA0038 and PA0043 tend towards major polymorphisms by the presence of two and three high-frequency alleles, respectively, in combination with a single low-frequency allele. Based on their evenly distributed allele frequencies, the multiallelic EST markers PA0002, PA0038 and PA0043 are expected to be highly sensitive in order to detect management impacts on the existing level of genetic diversity as well as to confirm genetic loads of natural populations (i.e. in-

breeding) and to indicate genotypic selection (see topic 3.3.), even in the case of small sample sizes. The respective markers therefore represent powerful tools for making genetic inventories of spruce populations and are warrant to replace such isoenzyme gene markers which exhibit inappropriate minor polymorphisms.

By employing the same 110 individuals of the Kranzberg test population, the results verified by EST markers are similar to the mean amount of genetic variation at 18 isoenzyme gene loci (see N_A values depicted in Figure 3). As furthermore indicated in Figure 3, the gene pool of six EST markers, however, reveals larger intrapopulation genetic variation than isoenzyme gene markers with respect to the mean observed heterozygosity (H_a), the mean number of genotypes per locus (N_G) as well as the mean diversity of alleles (V_A). The remarkable heterogeneity is mainly a consequence of the different types of polymorphisms observed: 80% of the analyzed 18 isoenzyme gene loci behaved or tended towards minor polymorphisms when the threshold for the individual predominant alleles was set to a minimum of 80% and 70% frequency, while only 50% of the assessed 6 EST markers tend towards minor polymorphisms as mentioned above (for detailed data see Schubert *et al.* 2001).

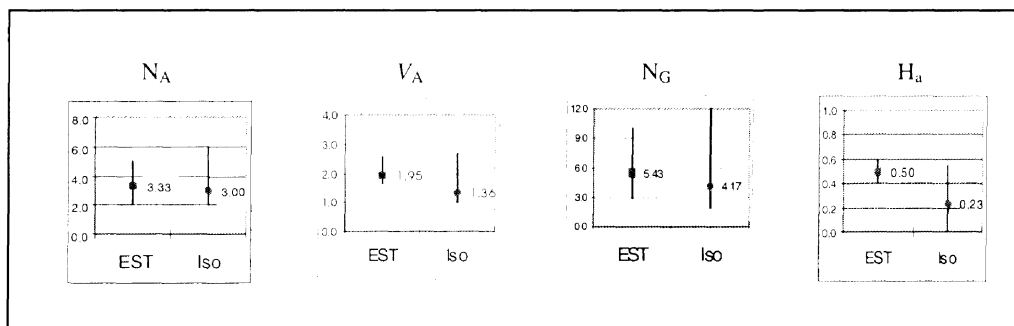


Figure 3. Minimum, maximum and mean values of genetic variation parameters summarized for 110 trees (bud samples) of the Kranzberg test population by analyzing 6 newly developed DNA markers (EST) and 18 isoenzyme markers (Iso). Based on the present number of alleles (N_A) and genotypes (N_G) per locus, diversity for alleles (V_A) and observed heterozygosity (H_a) were calculated according to Gregorius (1978, 1987) and Gregorius *et al.* 1986 (graphs taken from Schubert *et al.* 2001).

3.3 Genetic Characterization of Tolerant and Sensitive Subsets of Trees in Response to Abiotic and Biotic Stress

The long generation cycle of conifers substantially interferes with tree improvement because of the time required to reach trait expression and sexual maturity within the progenies of controlled crosses. Marker-assisted selection could therefore offer a more efficient alternate in realizing genetic gains with respect to stress-tolerant trees. Some of the newly developed EST markers of Norway spruce (see TABLE 1) are targeted to genes that are known in other species to be expressed in response to environmental stress. They are therefore expected to indicate genotypic selective effects, distinguishing between stress-tolerant and sensitive subsets of *Picea abies*.

Industrial air pollution has been recognized for long time as an important abiotic stress factor affecting survival of conifers. Since the 1950s, an extensive dieback of Norway spruce stands associated with SO₂-pollution have been observed in the higher elevations of the Ore Mountains (Saxony). Riegel (2001) has analyzed autovegetatively propagated SO₂-tolerant clone progenies, growing up at two neighbouring experimental plots in the locations Altenberg and Bärenfels (geographic distance approx. 10 km). The clone material has been originally selected in Germany, Poland and the Czech Republic for exhibiting long-term SO₂-resistance under field conditions (Tzschacksch 1981). During spring 1996, saplings suffered from the high-peak short-term SO₂-pollution which has been caused by the local climate situation in winter season 1995/1996 (Wolf 1999). For each location, 24 most responsive clones with respect to visible needle damage as well as 24 healthy clones were genotyped using the six new codominant EST markers mentioned. The genotypic distance (D) for marker PA0002, calculated between both sensitive and tolerant subsets of trees, is smaller than the genotypic distances existing between the sensitive and tolerant subsets of each plot as indicated by the cluster analysis given in the upper part of Figure 4. For the three genotypes which have been identified for marker PA0002, the frequency of the homozygous genotype AA is statistically significant increased in the pooled tolerant subsets of clones, representing both experimental plots, in comparison with both pooled sensitive subsets of clones (29% vs. 9% as shown in the lower part of Figure 4). Among the selected SO₂-resistent clones of Norway spruce, saplings with the homozygous genotype AA of marker PA0002 appear to be more tolerant under short-term episodes of SO₂-pollution. We therefore assume that the known cyclin gene itself or an unknown closely linked gene have an adaptive value under the abiotic stress conditions tested. Since not all individuals carrying the cyclin genotype AA are superior clones in terms of SO₂-pollution, it is likely that additional genomic loci, not yet discovered, are involved in the putative stress-defense mechanism.

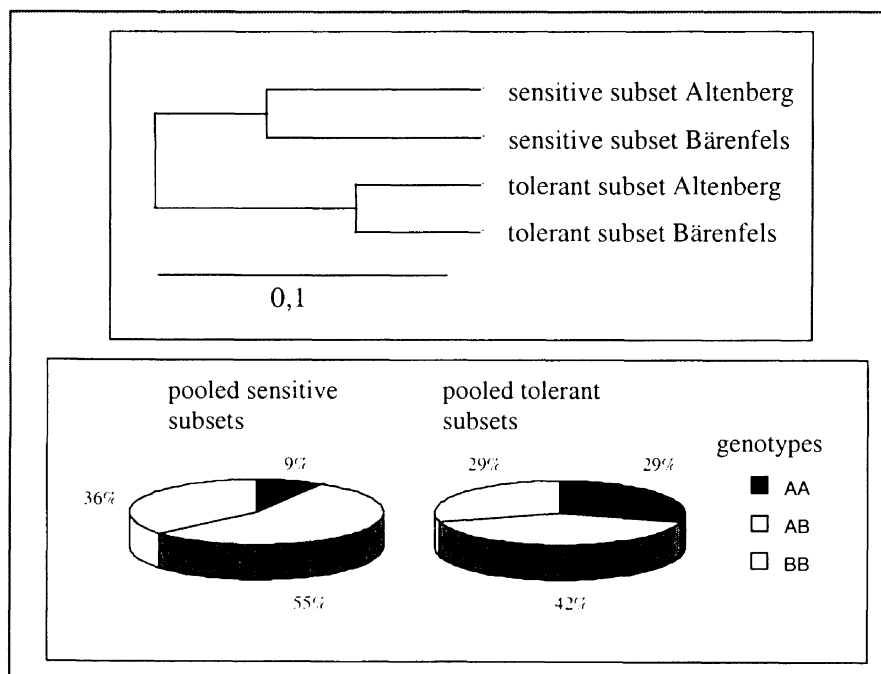


Figure 4. Upper part: UPGMA clustering of different subsets of SO₂-tolerant and sensitive spruce clones based on the genotypic distances D (Gregorius 1974) of EST marker PA0002. For each experimental plot at locations Altenberg and Bärenfels (Ore mountains), the 24 most sensitive clones with respect to needle damage were proved and compared with each healthy subset of 24 clones. Lower part: Frequency distributions for the present genotypes of marker PA0002 in both pooled subsets of sensitive and tolerant clones (data taken from Riegel 2001).

Based on this pilot study, data suggest that our new codominant EST markers, targeting stress-defense genes, are the first step towards marker-assisted selection with respect to complex abiotic stress. The majority of routinely used isoenzyme gene markers detects genetic variation in certain classes of water-soluble enzymes, which mainly represent the primary metabolism of plant cells. Since such house-keeping proteins are known to be constitutively expressed, isoenzyme gene markers are limited in their indicative value for selection caused by environmental stress (for comprehensive information see Müller-Starck and Schubert 2000).

Marker technologies furthermore offer powerful opportunities for characterizing the genetic basis of host pathogen interactions. For example, genomic mapping has been successfully used to identify a single gene in loblolly pine that determines resistance to fusiform rust disease (Wilcox *et al.* 1996). Oligonucleotide primers, recognizing genomic sequences of microorganisms, provide diagnostic probes ensuring the species-specific PCR-based detection of pathogens in infected conifers (e.g. Schulze and Bahnweg 1998). Since conifers are known to harbor pathogens even in the absence of visible disease symptoms (Bahnweg *et al.* 2000), reliable diagnostics and quantification methods are required to distinguish between susceptible and tolerant host trees for disease control and breeding programs.

3.4 Markers for Quantitative Trait Loci

The discovery of an expanding number of DNA polymorphisms and its integration into linkage maps offers prospects to identify markers which are associated with quantitative trait loci (QTL) involved in economic traits. In such cases the marker locus and the QTL will not segregate independently so that differences in those marker types will be linked with different phenotypic traits (for detailed information see Kearsey 1998). QTL mapping in non-domesticated predominantly outcrossing conifers is, however, complicated by the presence of multiple alleles at QTL and marker loci as well as by the lack of double haploid lines and backcross progenies of relatively large size. Phenotypic variation is therefore being assessed in outbred pedigrees, representing full-sib or half-sib families.

Based on RFLP markers, five QTL-associated genomic regions have been detected in loblolly pine influencing wood specific gravity (Groover *et al.* 1994). RAPD markers linked to QTLs of height, stem straightness, branch angle, and wood specific gravity have been identified in interspecific hybrid families of European and Japanese larches, involving a factorial mating design (Arcade *et al.* 1996). Moreover, QTLs were mapped for height growth components in maritime pine seedlings by RAPD markers, suggesting that the activity of the apical meristem and the subapical system is under the control of separate genetic mechanisms (Plomion *et al.* 1996). The genotype x environment interaction has been assessed for several growth traits in loblolly pine (McKeand *et al.* 1997). For this species, there are reports on QTLs with influence on annual-height- and diameter-increment growth (Kaya *et al.* 1999). Based on a new statistical method, QTLs controlling brown cone number, tree diameter, and branch quality have been estimated for radiata pine (Kao *et al.* 1999). Four QTL markers were also reported for timing of bud set and seven for frost hardiness in Scots pine, following the analysis of seeds collected from a single open-pollinated F₁ tree (Hurme *et al.* 2000). AFLPs are under investigation to discover QTLs for selected wood properties in Norway spruce (Markussen *et al.* 1999, 2001).

Several statistical approaches were developed for the identification of QTLs, but the main problems encountered with the application of QTLs reflect that heritabilities for most traits of interest are generally less than 50% and the phenotypic variation occurring in the test population represents only a small amount of the available within-species difference. Consequently, QTLs are known to have large confidence intervals when tested in different environments and genetic backgrounds. Nevertheless, QTL markers permit accelerated marker-assisted selection and are particularly suitable for forest trees where the assessment of economically important traits is usually unfeasible during the early stages of ontogenesis.

4. CONCLUSIONS

Genetic diversity is the basis for adaptation, evolution and survival of non-domesticated species under complex and unpredictable environmental conditions. Conservation of the naturally existing genetic diversity together with a sustainable management and the cultivation of well-adapted provenances are therefore providing the classical tools to contribute efficiently to ecologically stable forest tree populations and economic benefits with respect to specified population types and site conditions. Depending on future progress in genome analysis and molecular marker design, however, tree improvement is expected to speed up at an accelerated rate, even in the case of the slowly growing conifers which have not yet intensively managed by breeding programs. Once developed, marker-assisted selection assays including QTLs will become a great potential in breeding practices. Moreover, molecular markers offer new opportunities for establishing certified seed-lots and provenances to guarantee genetic quality under the conditions of free markets.

Nowadays, increased growth rates and additional genetic gain will be realized in the forest sector by silvicultural practices (Jain *et al.* 1995) and by the genetic engineering of such traits (e.g. herbicide tolerance, pathogen resistance, reduced lignin content; for recent survey see Kopriva and Renneberg 2000) which are not present in current breeding populations. To introduce foreign traits in conifers, genetic transformation has been established using *Agrobacterium* spp. and particle bombardment (Walter 1999). We argue that these non-conventional tools likely interfere with the necessity of maintaining high levels of genetic variability in long-lived forest tree populations.

Acknowledgements

Our research has been funded by grants of the European Community (Bio 4 CT96-0706) and the Deutsche Forschungsgemeinschaft (SFB 607). We wish to thank Ricardo Riegel (Freising, Germany) for providing *Figure 1* and *4*, and for processing *Figure 2*, as well as Christoph Sperisen (Birmensdorf, Switzerland) for providing the graphical base of *Figure 2*.

5. REFERENCES

- Ahuja, M.R., Devey, M.E., Groover, A.T., Jermstad, K.D. and Neale, D.B. (1994) Mapped probes from loblolly pine can be used for restriction fragment length polymorphism mapping in other conifers, *Theor. Appl. Genet.* **88**, 279-282.
- Allendorf, W. and Leary, R.F. (1986) Heterozygosity and fitness in natural populations of animals, in M.E. Soule (ed.), *Conservation biology: the science of scarcity and diversity*, Sinauer Associates, Sunderland, MA, pp. 57-76.
- Arcade, A., Rampant, P.F., Guerroue, B.L., Paques, L.E. and Prat, D. (1996) Quantitative traits and genetic markers: analysis of a factorial mating design in larch, in M.R. Ahuja (ed), *Somatic cell genetics and molecular genetics of trees*, Kluwer Academic Publishers, Netherlands, pp. 211-216.
- Bahnweg, G., Schubert, R., Kehr, R.D., Müller-Starck, G., Heller, W., Langebartels, C. and Sandermann, H. (2000), Controlled inoculation of Norway spruce (*Picea abies*) with *Sirococcus conigenus*: PCR-based quantification of the pathogen in host tissues and infection-related increase of phenolic metabolites, *Trees* **14**, 435-441.
- Baradat, P., Marpeau, A. and Walter, J. (1991) Terpene markers, in G. Müller-Starck and M. Ziehe (eds.), *Genetic variation in European populations of forest trees*, Sauerlaender's, Frankfurt am Main, pp. 40-66.
- Binelli, G. and Bucci, G. (1994) A genetic linkage map of *Picea abies* Karst., based on RAPD markers, as a tool in population genetics, *Theor. Appl. Genet.* **88**, 283-288.
- Bobola, M.S., Eckert, R.T. and Klein, A.S. (1992) Restriction fragment variation in the nuclear ribosomal DNA repeat unit within and between *Picea rubens* and *Picea mariana*, *Can. J. For. Res.* **22**, 255-263.
- Bobola, M.S., Eckert, R.T., Klein, A.S., Stapelfeldt, D.E.S. and Guenette, D. (1996) Using nuclear and organelle DNA markers to discriminate among *Picea rubens*, *Picea mariana*, and their hybrids, *Can. J. For. Res.* **26**, 433-443.
- Boudet, A.M., Lapierre, C. and Grima-Pettenati, J. (1995) Biochemistry and molecular biology of lignification, *New Phytol.* **129**, 203-236.
- Brandes, A., Heslop-Harrison, J.S., Kamm, A., Kubis, S. Doudrick, R.L. and Schmidt T. (1997) Comparative analysis of the chromosomal and genomic organization of *Tyl*-copialike retrotransposons in pteridophytes, gymnosperms and angiosperms, *Plant Mol. Biol.* **33**, 11-21.
- Bucci, G. and Vendramin, G.G. (2000) Delineation of genetic zones in the European Norway spruce natural range: preliminary evidence, *Mol. Ecology* **9**, 923-934.
- Burban, C., Petit, R.J., Carcreff, E. and Jactel, H. (1999) Rangwide variation of the maritime pine bast scale *matsucoccus feytaudi* duc. (Homoptera: matsucoccidae) in relation to the genetic structure of its host, *Mol. Ecology* **8**, 1593-1602.
- Carlson, J.E., Tulsieram, L.K., Glaubitz, J.C., Luk, V.W.K., Kauffeldt, C. and Rudlege, R. (1991) Segregation of random amplified DNA markers in F₁ progeny of conifers, *Theor. Appl. Genet.* **83**, 194-200.
- Cervera, M.T., Remington, D., Frigerio, J.M., Storme, V., Ivens, B., Boerjan, W. and Plo-mion, C. (2000) Improved AFLP analysis of tree species, *Can. J. For. Res.* **30**, 1608-1616.
- DeVerno, L.L. and Mosseler, A. (1997) Genetic variation in red pine (*Pinus resinosa*) revealed by RAPD and RAPD-RFLP analysis, *Can. J. For. Res.* **27**, 1316-1320.
- Devey, M.E., Jermstad, K.D., Tauer, C.G. and Neale, D.B. (1991) Inheritance of RFLP loci in loblolly pine three-generation pedigree, *Theor. Appl. Genet.* **83**, 238-242.
- Echt, C.S., May-Marquardt, P., Hseih, H. and Zahorchak, R. (1996) Characterization of microsatellite markers in eastern white pine, *Genome* **39**, 1102-1108.

- Echt, C.S., DeVerno, L.L., Anzidei, M. and Vendramin, G.G. (1998) Chloroplast microsatellites reveal population genetic diversity in red pine, *Pinus resinosa* Ait., *Mol. Ecology* **7**, 307-316.
- Elsik, C.G., Minihan, V.T., Hall, S.E., Scarpa, A.M. and Williams, C.G. (2000) Low-copy microsatellite markers for *Pinus taeda* L., *Genome* **43**, 550-555.
- FAO (1997) State of the World's Forests 1997, Rome.
- FAO (1999) State of the World's Forests 1999, Part I: Situation and Prospects for Forest Conservation and Development, Rome.
- Finkeldey, R. and Gregorius, H.R. (1991) Genetic resources: selection criteria and design, in Z.S. Kim, H.H. Hattemer (eds.), *Conservation and manipulation of genetic resources in forestry*, Seoul, Kwan Moon Kag Publishing Co, pp. 322-347.
- Govindaraju, R.D. and Cullis, C.A. (1991) Modulation of genome size in plants: the influence of breeding system and neighbourhood size, *Evol. Trends Plants* **5**, 43-51.
- Gregorius, H.R. (1974) Zur Konzeption der genetischen Abstandsmessung: Genetischer Abstand zwischen Populationen, *Silvae Genetica* **23**, 22-27.
- Gregorius, H.R. (1978) The concept of genetic diversity and its formal relationship to heterozygosity and genetic distance, *Math. Bioscience* **41**, 253-271.
- Gregorius, H.R. (1987) The relationship between the concepts of genetic diversity and differentiation, *Theor. Appl. Genet.* **74**, 397-401.
- Gregorius, H.R. (1991) Gene conservation and the preservation of adaptability, in A. Seitz, V. Loeschke (eds.), *Species Conservation: A Population-Biological Approach*, Basel, Birkhäuser Verlag, pp. 31-47.
- Gregorius, H.R., Krauhausen, J. and Müller-Starck, G. (1986) Spatial and temporal genetic differentiation among the seeds in a stand of *Fagus sylvatica* L., *Heredity* **57**, 255-262.
- Grima-Pettenati, J. and Goffner, D. (1999) Lignin genetic engineering revisited, *Plant Sci.* **145**, 51-65.
- Groover, A., Devey, M., Fiddler, T., Lee, J., Megraw, R., Mitchel-Olds, T., Sherman, B., Vujcic, S., Williams, C. and Neale, D. (1994) Identification of quantitative trait loci influencing wood specific gravity in an outbred pedigree of loblolly pine, *Genetics* **138**, 1293-1300.
- Guo, S.W. and Thompson, E.A. (1992) Performing the exact test of Hardy-Weinberg proportion for multiple alleles, *Biometrics* **48**, 361-372.
- Hamrick, J.L. and Godt, M.J.W. (1989) Allozyme diversity in plant species, in A.H.D. Brown, M.T. Clegg, A.L. Kahler and B.S. Weir (eds.), *Plant Population Genetics, Breeding and Genetic Resources*, Sinauer Associates, Sunderland, MA., pp. 77-104.
- Häger, K.P. and Fischer, H. (1999) Molecular phylogenies and structural diversification of gymnosperm and angiosperm storage globulines, in: R. Casey, P.R. Shewry (eds.), *Seed proteins*, Kluwer Academic Publishers, pp. 499-515.
- Hurme, P., Siilanjää, M.J., Arjas, E., Repo, T. and Savolainen, O. (2000) Genetic basis of climatic adaptation in scots pine by bayesian quantitative trait locus analysis, *Genetics* **156**, 1309-1322.
- Isabel, N., Tremblay, L., Michaud, M., Tremblay, F.M. and Bousquet, J. (1993) RAPDs as an aid to evaluate the genetic integrity of somatic embryogenesis-derived populations of *Picea mariana* (Mill.) B.S.P., *Theor. Appl. Gen.* **86**, 81-87.
- Jain, S.M., Gupta, P.K. and Newton, R.J. (1995) *Somatic embryogenesis in woody plants, Vol. 3. Gymnosperms*, Kluwer Academic Publishers, Dordrecht.
- Kao, C.H., Zeng, Z.B. and Teasdale, R.D. (1999) Multiple interval mapping for quantitative trait loci, *Genetics* **152**, 1203-1216.
- Kaya, Z., Sewell, M.M. and Neale, D.B. (1999) Identification of quantitative trait loci influencing annual height- and diameter-increment growth in loblolly pine (*Pinus taeda* L.), *Theor. Appl. Genet.* **98**, 586-592.
- Kearsey, M.J. (1998) The principles of QTL analysis (a minimal mathematics approach), *J. Exp. Bot.* **49**, 1619-1623.

- Kinlaw, C.S. and Neale, D.B. (1997) Complex gene families in pine genomes, *Trends in Plant Sci.* **2**, 356-359.
- Kopriva, S. and Rennenberg, H. (2000) Forest biotechnology and environment, *Bundesgesundheitsblatt, Gesundheitsforschung-Gesundheitsschutz* **43**, 121-125.
- Kvarnheden, A., Albert, V.A. and Engström, P. (1998) Molecular evolution of *cdc2* pseudo-genes in spruce (*Picea*), *Plant Mol. Biol.* **36**, 767-774.
- Lagercrantz, U. and Ryman, N. (1990) Genetic structure of Norway spruce (*Picea abies*): concordance of morphological and allozymic variation, *Evolution* **44**, 38-53.
- Latta, R.G. and Mitton, J.B. (1997) A comparison of population differentiation across four classes of gene marker in limber pine (*Pinus flexilis* James), *Genetics* **146**, 1153-63.
- Lian, C., Miwa, M., and Hogetsu, T. (2000) Isolation and characterization of microsatellite loci from Japanese red pine *Pinus densiflora*, *Mol. Ecology* **9**, 1186-1188.
- Lexer, C., Heinze, B., Gerber, S., Macalka-Kampfer, S., Steinkellner, H., Kremer, A. and Glössl, J. (2000) Microsatellite analysis of maternal half-sib families of *Quercus robur*, pedunculate oak: II. Inferring the number of pollen donors from the offspring, *Theor. Appl. Genet.* **100**, 858-865.
- MacKay, J.J., Liu, W.W., Whetten, R., Sederoff, R.R. and O'Malley, D.M. (1995) Genetic analysis of cinnamyl alcohol dehydrogenase in loblolly pine: single gene inheritance, molecular characterization and evolution, *Mol. Gen. Genet.* **247**, 537-545.
- Markussen, T., Tusch, A., Stephan, R. and Fladung, M. (1999) QTL mapping of selected wood properties in Norway spruce (*Picea abies* (L.) Karst.), in International Congress: Application of biotechnology to forest genetics, Vitoria-Gasteiz, Spain, September 22-25, Book of Abstracts, pp. 157-159.
- Markussen, T., Tusch, A., Stephan, B.R. and Fladung, M. (2001) Identification of molecular markers for selected wood properties of Norway spruce (*Picea abies* (L.) Karst.), in International Congress: Wood, breeding, biotechnology and industrial expectations, Bordeaux, France, June 11-14, Book of Abstracts, p. 56.
- Marshall, K.A. and Neale, D.B. (1992) The inheritance of mitochondrial DNA in Douglas-fir (*Pseudotsuga menziesii*), *Can. J. For. Res.* **22**, 73-75.
- McKeand, S.E., Eriksson, G. and Roberds, J.H. (1997) Genotype by environment interaction for index traits that combine growth and wood density in loblolly pine, *Theor. Appl. Genet.* **94**, 1015-1022.
- Mitton, J.B. (1983) Conifers, in S.D. Tanksley and T.J. Orthon (eds.), *Isoenzymes in Plant Genetics and Breeding*, Part B, Elsevier, Amsterdam, New York, pp. 443-472.
- Mitton, J.B. and Grant, M.C. (1984) Associations among protein heterozygosity, growth rate, and developmental homeostasis, *Annu. Rev. Ecol. Syst.* **15**, 479-499.
- Müller-Starck, G. (1991) Survey of genetic variation as inferred from enzyme gene markers, in G. Müller-Starck and M. Ziehe (eds.), *Genetic variation in European populations of forest trees*, Sauerlaender's, Frankfurt am Main, pp. 20-37.
- Müller-Starck, G. and Ziehe, M. (1991) *Genetic Variation in European Populations of Forest Trees*, Sauerlaender's, Frankfurt am Main, pp. 271.
- Müller-Starck, G., Baradat, P. and Bergmann, F. (1992) Genetic variation within European tree species, *New Forests* **6**, 23-47.
- Müller-Starck, G. and Schubert, R. (2000) Genetic markers as a tool for bioindication in forest ecosystems, in A. Young, D. Boshier and T. Boyle (eds.), *Forest Conservation Genetics: Principles and Practice*, CSIRO Publishing, Collingwood, Australia, and CABI Publishing, Wallingford, UK, pp. 227-237.
- Niklas, K.J., Tiffney, B.H. and Knoll, A.H. (1983) Patterns in vascular land plant diversification, *Nature* **303**, 614-616.
- Paglia, G.P., Olivieri, A.M. and Morgante, M. (1998) Towards second-generation STS (sequence-tagged sites) linkage maps in conifers: a genetic map of Norway spruce (*Picea abies* K.), *Mol. Gen. Genet.* **258**, 466-478.

- Pearce, S.R., Stuart-Rogers, C., Knox, M.R., Kumar, A., Noel Ellis, T.H. and Flavell, A.J. (1999) Rapid isolation of plant *Ty1*-copia group retrotransposon LTR sequences for molecular marker studies, *Plant J.* **19**, 711-717.
- Perry, D.J. and Furnier, G.R. (1996) *Pinus banksiana* has at least seven expressed alcohol dehydrogenase genes in two linked groups, *Proc. Natl. Acad. Sci. USA* **93**, 13020-13023.
- Perry, D.J. and Bousquet, J. (1998) Sequence-tagged-site (STS) markers of arbitrary genes: development, characterization and analysis of linkage in black spruce, *Genetics* **149**, 1089-1098.
- Perry, D.J., Isabel, N., Bousquet, J. (1999) Sequence-tagged-site (STS) markers of arbitrary genes: the amount and nature of variation revealed in Norway spruce, *Heredity* **83**, 239-248.
- Pfeiffer, A., Olivieri, A.M. and Morgante, M. (1997) Identification and characterization of microsatellites in Norway spruce (*Picea abies* K.), *Genome* **40**, 411-419.
- Plomion, C., Durel, C.E. and O'Malley, D.M. (1996) Genetic dissection of height in maritime pine seedlings raised under accelerated growth conditions, *Theor. Appl. Genet.* **93**, 849-858.
- Powell, W., Machray, G.C. and Provan, J. (1996) Polymorphism revealed by simple sequence repeats, *Trends in Plant Sci.* **1**, 215-222.
- Prus-Glowacki, W. (1982) Immunochemical methods in analysis of forest tree proteins, *Silvae Fennica* **16**, 219-226.
- Remington, D.L. and O'Malley, D.M. (2000) Whole-genome characterization of embryonic stage inbreeding depression in a selfed loblolly pine family, *Genetics* **155**, 337-348.
- Riegel, R. (2001) Entwicklung molekulargenetischer Marker bei der Fichte [*Picea abies* (L.) Karst.] und deren Anwendung für genetische Erhebungen in umweltbelasteten Populationen, Thesis, Technical University of Munich, 97 Seiten.
- Scheepers, D., Eloy, M.C. and Briquet, M. (1997) Use of RAPD patterns for clone verification and in studying provenance relationships in Norway spruce (*Picea abies*), *Theor. Appl. Genet.* **94**, 480-485.
- Schubert, R., Müller-Starck, G., Sandermann, H., Ernst, D. and Häger, K.P. (1997) The molecular structure and evolutionary relationships of a 16.9 kDa heat shock protein from Norway spruce [*Picea abies* (L.) Karst.], *For. Genet.* **4**, 131-138.
- Schubert, R., Sperisen, C., Müller-Starck, G., La Scala, S., Ernst, D., Sandermann, H. and Häger, K.P. (1998) The cinnamyl alcohol dehydrogenase gene structure in *Picea abies* (L.) Karst.: genomic sequences, Southern hybridization, genetic analysis and phylogenetic relationships, *Trees* **12**, 453-463.
- Schubert, R., Riegel, R. and Müller-Starck, G. (2001) Development of EST-PCR markers and monitoring their intrapopulational genetic variation in *Picea abies* (L.) Karst., *Theoret. Appl. Genet.*, in press.
- Schulze, S. and Bahnweg, G. (1998) Identification of the genus *Armillaria* (Fr.: Fr.) Staude and *Heterobasidion annosum* (Fr.) Bref. in Norway spruce (*Picea abies* [L.] Karst.) and determination of clonal distribution of *A. ostoyae*-genotypes by molecular methods, *Forstw. Cbl.* **117**, 98-114.
- Scotti, I., Troggio, M., Soranzo, N., Vendramin, G.G. and G. Bucci (1998) A new set of PCR-based, locus-specific markers for *Picea abies* (L.) Karst., *Molecular Ecology* **7**, 783-792.
- Scotti, I., Magni, F., Fink, R., Powell, W., Binelli, G. and Hedley P.E. (2000a) Microsatellite repeats are not randomly distributed within Norway spruce (*Picea abies* K.) expressed sequences, *Genome* **43**, 41-46.
- Scotti, I., Vendramin, G.G., Matteotti, L.S., Scarponi, C., Sari-Gorla, M. and Binelli, G. (2000b) Postglacial recolonization routes for *Picea abies* K. in Italy as suggested by the analysis of sequence-characterized amplified (SCAR) markers, *Mol. Ecology* **9**, 699-708.
- Sperisen, C., Büchler, U., Gugerli, F., Matyas, G. and Geburek, T. (2001) Tandem repeats in plant mitochondrial genomes: application to the analysis of population differentiation in the conifer Norway spruce, *Mol. Ecology* **10**, 257-263.

- Stine, M. and Keathley, D.E. (1990) Paternal inheritance of plastids in Engelmann spruce x Blue spruce hybrids, *J. of Heredity* **81**, 443-446.
- Strauss, S.H., Hong, Y.P. and Hipkins, V.D. (1993) High levels of population differentiation for mitochondrial DNA haplotypes in *Pinus radiata*, *muricata*, and *attenuata*, *Theor. Appl. Genet.* **86**, 605-611.
- Sunnucks, P. (2000) Efficient genetic markers for population biology, *Tree* **15**, 199-203.
- Sutton, B.C.S., Flanagan, D.J., Gawley, J.R., Newton, C.H., Lester, D.T. and El-Kasaby, Y.A. (1991) Inheritance of chloroplast and mitochondrial DNA in *Picea* and composition of hybrids from introgression zones, *Theor. Appl. Genet.* **82**, 242-248.
- Temesgen, B., Brown, G.R., Harry, D.E., Kinlaw, C.S., Sewell, M.M. and Neale, D.B. (2001) Genetic mapping of expressed sequence tag polymorphism (ESTP) markers in loblolly pine (*Pinus taeda* L.), *Theor. Appl. Genet.* **102**, 664-675.
- The Arabidopsis Genome Initiative (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*, *Nature* **408**, 796-826.
- Tsumura, Y., Suyama, Y., Yoshimura, K., Shirato, N. and Mukai, Y. (1997) Sequence-tagged-sites (STS) of cDNA clones in *Cryptomeria japonica* and their evaluation as molecular markers in conifers, *Theor. Appl. Genet.* **94**, 764-772.
- Tulsieram, L.K., Glaubitz, J.C., Kiss, G. and Carlson, J.E. (1992) Single tree genetic linkage mapping in conifers using haploid DNA from megagametophytes, *Biotechnology* **10**, 686-690.
- Tzschacksch, O. (1981) Stand und Perspektiven der forstlichen Rauchresistenzzüchtung in der DDR, *Beiträge für die Forstwirtschaft* **15**, 134-137.
- Van de Ven, W.T.G. and McNicol, R.J. (1996) Microsatellites as DNA markers in Sitka spruce, *Theor. Appl. Genet.* **93**, 613-617.
- Vendramin, G.G., Anzidei, M., Madaghiele, A., Sperisen, C. and Bucci, G. (2000) Chloroplast microsatellite analysis reveals the presence of population subdivision in Norway spruce (*Picea abies* K.), *Genome* **43**, 68-78.
- Vendramin, G.G., Lelli, L., Rossi, P. and Morgante, M. (1996) A set of primers for the amplification of 20 chloroplast microsatellites in *Pinaceae*, *Mol. Ecology* **5**, 595-598.
- Vendramin, G.G. and Ziegenhagen, B. (1997) Characterisation and inheritance of polymorphic plastid microsatellites in *Abies*, *Genome* **40**, 857-864.
- Vos, P., Hogers, R., Bleeker, M., Reijans, M., Van der Lee, T., Hornes, M., Frijters, A., Pot, J., Peleman, J., Kuiper, M. and Zabeau, M. (1995) AFLP: a new technique for DNA fingerprinting, *Nucl. Acids Res.* **23**, 4407-4414.
- Wagner, D.B., Govindaraju, D.R., Yeatman, C.W. and Pitel, J.A. (1989) Paternal chloroplast DNA inheritance in a diallel cross of jack pine (*Pinus banksiana* Lamb.), *J. of Heredity* **80**, 483-485.
- Walter, C. (1999) Genetic transformation of forest tree species. State of the art and future challenges. in International Congress: Application of biotechnology to forest genetics, Vitoria-Gasteiz, Spain, September 22-25, Book of Abstracts, pp. 331-338.
- Wakasugi, T., Tsudzuki, J., Ito, S., Nakashima, K., Tsudzuki, T., Sugiura, M. (1994) Loss of all *ndh* genes as determined by sequencing the entire chloroplast genome of the black pine *Pinus thunbergii*, *Proc. Nat. Acad. Sci. USA* **91**, 9794-9798.
- White, G. and Powell, W. (1997) Isolation and characterization of microsatellite loci in *Swietenia humilis* (Meliaceae): an endangered tropical hardwood species. *Mol. Ecology* **6**, 851-860.
- Wilcox, P.L., Amerson, H.V., Kuhlman, E.G., Liu, B.H., O'Malley, D.M. and Sederoff, R.R. (1996) Detection of a major gene for resistance to fusiform rust disease in loblolly pine by genomic mapping, *Proc. Natl. Acad. Sci. USA* **30**, 3859-3864.
- Williams, J.G.K., Kubelik, A.R., Livak, K.J., Rafalski, J.A. and Tingey, S.V. (1990) DNA polymorphisms amplified by arbitrary primers are useful as genetic markers, *Nucleic Acids Res.* **18**, 6531-6535.

- Wolf, H. (1999) Effects of extreme SO₂ air pollution in winter 1995/96 on vitality and growth of SO₂-tolerant Norway spruce (*Picea abies* [L.] Karst.) clones in the Ore Mountains. in IUFRO Conference: Genetic response of forest systems to changing environmental conditions -Analysis and management-, University of Munich, Freising, Germany, September 12-17, Book of Abstracts, p. 19.
- Wolff, R.L., Deluc, L.G., Marpeau, A.M. and Comps, B. (1997) Chemotaxonomic differentiation of conifer families and genera based on the seed oil fatty acid composition: multivariate analyses, *Trees* **12**, 75-65.
- Yokoyama, S. and Harry, D.E. (1993) Molecular phylogeny and evolutionary rates of alcohol dehydrogenases in vertebrates and plants, *Mol. Biol. Evol.* **10**, 1215-1226.
- Young, A., Boshier, D. and Boyle, T. (2000) *Forest Conservation Genetics, Principles and Practice*, CSIRO Publishing, Collingwood, Australia and CABI Publishing, Wallingford, UK.
- Ziegenhagen, B., Guillemaut, P. and Scholz, F. (1993) A procedure for mini-preparations of genomic DNA from needles of Silver fir (*Abies alba* Mill.), *Plant Mol. Biol. Rep.* **11**, 117-121.
- Ziegenhagen, B. and Fladung, M. (1997) Variation in the *psbC* gene region of gymnosperms and angiosperms as detected by a single restriction site polymorphism, *Theor. Appl. Genet.* **94**, 1065-1071.
- Ziegenhagen, B., Scholz, F., Madaghiele, A., and Vendramin, G.G. (1998) Chloroplast microsatellites as markers for paternity analysis in *Abies alba*, *Can. J. For. Res.* **28**, 317-321.

6

DNA MARKERS AND HETEROSIS

PRABHAKAR K. RANJEKAR*, ARMAITY P. DAVIERWALA, VIDYA S. GUPTA

Plant Molecular Biology Unit, Biochemical Sciences Division, National Chemical Laboratory, Pashan, Pune 411008 (INDIA)
Tel: 91-20-5893034; Fax: 91-20-5884032

* Communicating author: Email: pkrr@ems.ncl.res.in

INTRODUCTION

The growth in human population every year all over the world has resulted in an increased demand for food crops. With decreasing resources such as arable land and water, the only suitable alternative is to increase the yield of crop plants. Biotechnology is looked as a potential tool to significantly strengthen and complement the current crop breeding programs. Developing high yielding and superior hybrids with the advent of biotechnology is one of the significant strategies to break the yield barriers.

(I) IMPORTANCE OF HETEROSIS / HYBRID VIGOR IN PLANT BREEDING

The basic requirement for producing hybrid crops is to obtain a hybrid with a sufficient level of heterosis to justify the higher cost of seed production. The term heterosis was first coined by Shull (1914), for the special stimulus of heterozygosis upon cell division, growth and other physiological activities of an organism. It is basically a new term for hybrid vigor, which had earlier been described by Shull (1908, 1911) and East (1908). Shull (1914) defined heterosis as the increase in size, yield and vigor. This definition was later

modified by evolutionists, to include heterosis in survival, ie. adaptive, selective and reproductive advantage (Dobzhansky, 1950).

Heterosis or hybrid vigor is the phenomenon in which the performance of an F1 plant, generated by crossing of two genetically different individuals, is superior to that of the mid / better parent. The phenomenon of heterosis has been successfully exploited in commercial hybrids by many plant breeders, to enhance the productivity of various crops, such as rice, maize, sorghum, pearl millet and cotton, and horticultural plants, and the effects of heterosis have been quantified in a wide variety of plant studies. However, inspite of several attempts, there is as yet no satisfactory explanation for the biological basis or the genetic cause of heterosis. No other biological phenomenon is probably shrouded in so much mystery as heterosis, with respect to the probable cause of its manifestation. As Sprague has (1983) stated, "the casual factors for heterosis at the physiological and biochemical level are today almost as obscure as they were 50 years ago". Understanding the phenomenon of heterosis will help in creating new and more productive hybrids and to generate useful genetic variability in the progeny lines, prior to the selection of superior lines after field evaluation of their phenotype. Crosses that give rise to heterotic F1s and superior F2 lines usually result in the most desirable genetic variation for the fastest progress to be attained through selection (Fasoulas, 1988).

(II) GENETIC BASIS OF HETEROSIS

In the early years, dominance and overdominance hypothesis have been put forward to explain the genetic basis of heterosis. In the dominance hypothesis put forth by Davenport (1908), Bruce (1910) and Keeble and Pellew (1910), and later elaborated by Jones (1917), it is suggested that genes governing vigor are beneficial and dominant in nature while recessives are unfavourable and have a detrimental effect on heterosis. Also, in heterozygous state the deleterious effects of the recessive alleles are masked by their dominant alleles. According to the overdominance hypothesis, postulated by Shull (1908; 1911) and East (1908), heterozygotes are superior to both the relevant homozygotes at least at some of the loci. Pseudo-overdominance (Crow, 1952) (i.e. near-by loci at which alleles having dominant or partially dominant advantageous effects are in repulsion phase linkage) is an extension of the hypothesis of dominance of linked factors put forth by Jones (1917). In pseudo-overdominance, the linked block of two or more loci tends to respond

as a single locus and the effect is nearly impossible to distinguish from true overdominance particularly in plants, since the evidence for true overdominance is very limited.

Recent reports contrary to previous assumptions have shown that interaction of non-alleles (epistasis) may be an important factor in the type of heterosis found in many maize single crosses (Otsuka, *et al*, 1972; Stuber, *et al*, 1973) and more recently in rice (Yu *et al*, 1997). Minville (1987) has further suggested that dominance may not be necessary for heterosis with the presence of epistasis. Schnell and Cockerham (1992) have reviewed the influence exerted on heterosis by multiplicative effects between genes (a specific type of epistasis). Heterosis in quantitative traits may arise from the action and interaction of polygenes and such interactions could be inter-allelic, intra-allelic or non-allelic. The available techniques have failed to precisely differentiate among such interactions and hence, no genetic theory can entirely explain the expression of heterosis. Several studies have been carried out in maize to determine which of the hypothesis could best explain heterosis (Sprague and Miller, 1950; Jones, 1952; Russell, and Eberhart, 1970; Russell, 1971; Russell, *et al*, 1973; Walejko and Russell, 1977; Lamkey and Smith, 1987; Stuber *et al*, 1992). However, all these studies could not conclusively explain the genetic basis of heterosis and it is possible that the two hypothesis are not mutually exclusive and that heterosis could result from the action of both kinds of genes (Tsaftaris, 1995).

All the early work on heterosis has been done using conventional breeding methods. In order to develop new elite commercial hybrids by conventional breeding strategies, several crosses need to be performed using many different genotypes and the progeny need to be screened for superior F1 performance and strong heterosis. Evaluation of hybrids for heterosis or combining ability by field trials is not only laborious and time consuming but also very expensive for any breeding program. If a simple, quick, efficient, inexpensive and reliable method could be used to predict heterosis prior to expensive field trials, much of the field work associated with performing a large number of crosses and their field evaluation would be eliminated and hybrid programs would be accelerated.

(III) EXPLOITING THE UTILITY OF MOLECULAR MARKERS FOR ANALYSIS OF HETEROSIS

Molecular genetic markers are heritable entities that are associated with economically important traits and can be used by plant breeders as selection tools (Beckmann and Soller, 1983; Darvasi, and Soller, 1994). These markers which represent the genetic variation, can be used to estimate the degree of relatedness between different genotypes and can subsequently be used to predict which crosses might produce new, superior and desirable gene combinations. If markers are available for the genes of interest, it would be easily possible to detect new recombinations between these genes and thus help to select individuals which are genetically superior (Paterson, *et al*, 1991).

Molecular markers are discreet, nondeleterious and unaffected by environmental factors and are, therefore, superior to morphological markers. They are basically of two types, biochemical markers and DNA markers.

(A) Biochemical markers

Biochemical markers reveal polymorphism at the protein level and the most commonly used protein markers are isozymes, which are variant forms of the same enzyme (Vodenicharova, 1989). Allelic diversity at isozyme loci has been correlated with grain yield of single-cross hybrids derived from commercially inbred maize lines (Hunter and Kannenberg, 1971; Heidrich-Sobrinho and Cordeiro, 1975; Gonella and Peterson, 1978; Hadjinov *et al*, 1982; Tsiftaris and Efthimiadis, 1987). It has been observed in these studies that the correlations between isozyme diversity and specific combining ability are low and non-significant, when 15 inbred lines have been analyzed at a small number of isozyme loci (≤ 11). Smith and Smith (1989) have also observed a low correlation value (r) when 100 maize hybrids derived from 37 parental lines have been evaluated at 30 isozyme loci. They have obtained an r^2 value of 0.36 when F1 yield was plotted against isozyme allelic diversity. However, Tsiftaris (1995) has reported a small but significant correlation ($r=0.23$) on analysis of 16 maize inbreds at 47 isozyme loci. All these studies prove that diversity at isozyme loci cannot be used as a true predictor of hybrid performance. The reason for this could be that only a limited number of isozyme markers are available in any crop species and, therefore, the number of isozyme loci assayed in most studies would effectively cover only a small fraction of the entire genome. Also these markers are subject to post-translational modifications (Staub *et al*, 1982), thus making them unsuitable for this analysis.

(B) DNA markers

With the advent of DNA markers, large number of loci could be analyzed, thus providing wider genome coverage, which is not possible using isozyme markers. These markers reveal polymorphism at the DNA level and can be basically classified into two categories, hybridization-based markers and PCR-based markers. Till date several types of DNA markers have been made available for use in various agricultural applications (Joshi *et al*, 1999).

The availability of genetic linkage maps for different crops has made it possible to use these molecular markers to tag agronomically important genes for use in hybrid breeding programs. Molecular markers can be used in hybrid programs for quantifying relatedness in parental stocks, verification of pedigree records, assigning inbreds to heterotic groups, prediction of heterosis and hybrid performance, tagging genes and quantitative trait loci (QTLs) for heterosis and understanding the basis of heterosis.

(IV) SELECTION OF BETTER COMBINING PARENTS USING MOLECULAR MARKERS

Information on genetic diversity in a crop species is important for selection of suitable parental genotypes. Predicting hybrid performance and heterosis has always been the primary objective in all hybrid crop breeding programs (Hallauer and Miranda, 1988). In crops such as rice and maize, where hybrids are commercially important, it is imperative to exploit the heterosis among genetically divergent germplasm because the level of genetic diversity has been proposed as a possible predictor of F1 performance and heterosis. It has been observed that rice hybrids showing strong heterosis were usually developed from parental lines diverse in relatedness, ecotype, geographic origin, etc (Lin and Yuan, 1980; Yuan, 1985; Yuan and Cheng, 1986). Experimental studies (Moll *et al*, 1965) have illustrated that mid-parent heterosis could be related to genetic divergence, a relationship that has been supported by quantitative genetics theory (Falconer, 1981). In maize, heterotic groups have been established by relating heterosis observed in crosses to the origin of parents, morphological traits, geographic origin, and, to a limited extent, isozyme diversity (Brar *et al*, 1994). Deng and Wang (1984) have demonstrated that dominant, complementary bands in the isozyme spectra are correlated with magnitude of heterosis. However, Peng *et al* (1988) have

found no correlation between isozyme diversity and hybrid performance. Stuber (1989) has reviewed literature on the use of isozymes as possible predictors of hybrid performance in maize and other crops. Distances computed using data generated at isozyme loci could be significantly correlated to heterosis in some cases. However, the correlations have been generally too low for the distances to be of practical predictive value. A possible explanation for the limited informativeness provided by isozymes, is the relatively few polymorphic marker loci that are available in the elite germplasm. Investigators are, therefore, now trying to correlate genetic diversity as revealed by DNA markers, to predict hybrid performance in various hybrid breeding programs.

(A) Molecular markers as predictors of hybrid performance in maize

RFLPs have been used in several crops for prediction of hybrid performance. When RFLPs have been used to examine the relationship between RFLP-based genetic distance and single cross grain yield in maize inbred lines, relatively small positive correlations have been obtained between parental genetic distance and F1 yield and heterosis (Lee *et al*, 1989; Godshalk *et al*, 1990; Melchinger *et al*, 1990a; 1991; Dudley *et al*, 1991; Boppenmaier *et al*, 1992). However, when forage yield has been measured instead of grain yield, RFLP genetic distance has been found to be a useful predictor of forage yield in maize inbred lines (Melchinger *et al*, 1992). In all these studies, the number of lines involved has been relatively small (<20). Stuber *et al* (1992) have observed that, when the number of maize inbred lines is increased (230 lines have been analyzed), a highly significant relationship has been obtained between parental heterozygosity and hybrid yield. Smith *et al* (1991) have demonstrated that RFLP diversity among maize hybrids correlates with genetic distance based on multigenic traits such as pedigree, F1 yield and yield heterosis. Several studies using RFLP based estimates of genetic similarity among elite maize inbreds have demonstrated the utility of DNA markers for placing the lines in their respective heterotic groups (Lee *et al*, 1989; Godshalk *et al*, 1990; Melchinger *et al*, 1991; Dudley *et al*, 1991; Livini *et al*, 1992; Messmer *et al*, 1993) rather than in predicting heterotic performance of single crosses between unrelated lines (Melchinger *et al*, 1990a).

Studies on maize (Godshalk *et al*, 1990, Melchinger *et al*, 1990a, Boppenmaier *et al*, 1993) and oat (Moser *et al*, 1994) have revealed that

molecular marker-based genetic distances could be a predictor of hybrid performance only in those crosses where parents belong to the same heterotic group, and cannot be extended to crosses between different heterotic groups. Thus, the correlation between genetic distance and hybrid value increases with increasing relatedness in the germplasm under study (Stuber, *et al*, 1992). AFLP markers have also been used to study the correlation between genetic diversity and hybrid performance in maize (Ajmone Marsan *et al*, 1998). Despite the inconsistency in the correlations, results from all these studies indicate that genetic distances based on marker genotypes are in close agreement with pedigree information and can unambiguously resolve lines into their respective heterotic groups. RAPD-based genetic distances have been found to correlate consistently with grain yield in maize (Larza *et al*, 1997). In wheat, no correlation has been observed between marker-based genetic distance and hybrid performance (Martin *et al*, 1995; Barbosa-Neto *et al*, 1996). In soybean and oilseed rape, genetic distance has not been significantly and consistently correlated with heterosis and, therefore, has not been useful to make any predictions (Diers *et al*, 1996; Cerna *et al*, 1997).

(B) Studies of hybrid performance and heterosis in rice

The world production of hybrid rice parallels the role of hybrid corn in many ways. Hybrid rice has contributed significantly to the dramatic increase in the rice production of the world (Yuan, 1992a). Extensive breeding efforts to develop superior performing elite rice hybrids, have resulted in the release of several hybrid combinations in China and other Asian countries during the last 20 years. Consequently, a very large area is now being planted with hybrid rice. In China alone, it has been estimated that about 17 million ha are planted with F1 hybrids annually, which is approximately 55% of the total rice area under cultivation there.

Very few studies have been reported estimating the amount of heterosis in varietal crosses and demonstrating the performance of hybrids (Rice Research Group, 1978; Virmani *et al*, 1982; Young and Virmani, 1990; Virmani *et al*, 1991). Two measures of heterozygosity used to calculate the correlations of heterozygosity with performance and heterosis, are general heterozygosity, which is the heterozygosity at all the marker loci and specific heterozygosity, which is the heterozygosity calculated on the basis of positive markers for each trait. Using RFLPs and STMS markers, low correlation has

been obtained between heterozygosity and F1 performance, while high correlations have been obtained between specific heterozygosity and mid-parent heterosis (Zhang *et al*, 1994a; 1995), indicating that specific heterozygosity may find practical applications for prediction of heterosis. These results are in agreement with previous studies in corn (Stuber *et al*, 1992). Thus, in reality, it may be much more practical to predict heterosis on the basis of a small number of informative markers rather than to use a large number of markers covering the entire genome corroborating the suggestion from theoretical calculations (Bernardo, 1992, Charcosset, *et al*, 1991). Molecular analysis of eight lines, representing a major portion of the elite rice germplasm used in hybrid rice programs in China, has resulted into three well separated groups which essentially agreed with the available pedigree information (Zhang *et al*, 1995). However, if such correlations can be confirmed using a larger sample size rather than just eight parental lines used in these analyses, it can help in planning the most productive crosses in hybrid breeding programs.

Zhang *et al* (1996) have observed varying degrees of correlation between molecular marker heterozygosity and heterosis in indica and japonica rice. When RFLP and STMS markers have been used to analyze U.S. Southern long grain rice, F1 heterozygosity has been highly correlated with rough rice yield and head rice yield, and significantly correlated with heterosis of these two traits, suggesting that molecular markers may be partially useful for predicting the performance of inter-subspecific hybrids (Saghai Maroof *et al*, 1997). However, studies including more indica-tropical japonica hybrids are needed in order to establish the relationship between marker heterozygosity and hybrid performance in such inter-subspecific crosses and to assess the feasibility of exploiting the heterosis between indica and tropical japonica in hybrid rice breeding programs. Xiao *et al* (1996) have observed a significant positive correlation between heterosis in F1 hybrid and genetic distance based on RAPD and STMS markers in ten elite inbred lines consisting of four japonicas and six indicas, widely used in the hybrid rice research program. They have concluded that genetic distance measures based on these two molecular marker types, may be useful for predicting yield potential and heterosis of intra-subspecific hybrids, but not of inter-subspecies hybrids. This suggests that parental genotyping based on RAPD and STMS markers could be useful in reducing the field work associated with making crosses and hybrid field testing when attempting to identify intra-subspecies hybrids possessing

high yield potential as targets for transfer of key genes in rice. All these studies clearly demonstrate that the extent and modes of correlation between marker heterozygosity and hybrid performance in rice are also dependent on the genetic material used in the studies as suggested by theoretical calculations (Benardo, 1992) and as reviewed in corn (Melchinger, 1993).

(C) Correlation of genetic distance with hybrid performance - Indian scenario

In our laboratory, attempts have been made to correlate genetic distance with hybrid performance and heterosis in crops like rice, chickpea and pearl millet. In chickpea, eight elite parental lines and a diallel set of 28 crosses; obtained by crossing these parental lines, have been analyzed using RAPDs and microsatellites. Although no linear correlation is obtained between genetic diversity and heterosis, significant heterosis has resulted for hybrids obtained by crossing parents from two different subgroups, while hybrids obtained by crossing parents from the same group has resulted into poor heterosis. (Sant *et al*, 1999). Although molecular marker based genetic distance could not be linearly correlated to heterosis, two heterotic groups have been identified on the basis of the general marker heterozygosity. Most such studies to correlate molecular marker genetic diversity with heterosis have been carried out using diallel crosses. Studies using the microsatellite (GATA)₄ and RAPD markers (Chowdari *et al*, 1998a), on five cytoplasmic male sterile (CMS) lines and seven restorer lines of pearl millet have revealed that correlation of genetic distance with hybrid performance and heterosis is significant only for days-to-50% flowering trait, while for the trait of ear length and ear width, heterosis over the mid-parent and the better parent are significantly correlated with genetic distance (Chowdari *et al*, 1998b). Similarly a three line hybrid system, involving two cytoplasmic male sterile lines and fourteen restorer lines of rice, widely, grown in western regions of Maharashtra, India, has been used for correlation studies using random markers such as RAPD and ISSR and specific markers such as RFLP and STMS. Random heterozygosity markers could not provide any significant positive correlation indicating that they are not of much predictive value for hybrid performance and heterosis. However, specific heterozygosity markers have produced positive correlation in a number of traits especially for hybrid performance, when the effect was measured collectively rather than individually (Joshi *et al*, 2001).

In conclusion, a general consensus view derived from all these reports is that random marker analysis can be used to assign inbreds to heterotic groups and to quantify genetic similarities between related lines thus providing useful means for germplasm characterization. However, all the studies on different crop plants suggest the use of specific heterozygosity markers, significantly associated or linked with the trait responsible for heterotic yield, rather than general heterozygosity markers, spanning the entire genome and showing no significant association with the trait, for prediction of F1 hybrid performance and heterosis (Melchinger *et al*, 1990a; 1990b; Boppenmaier *et al*, 1993; Moser and Lee, 1994; Zhang *et al*, 1994a; 1995; 1996; Joshi *et al*, 2001). The low correlation between genetic distance and hybrid performance could also result due to inadequate genome coverage and different levels of dominance among hybrids (Melchinger *et al*, 1990b; Bernardo, 1992). Charcosset *et al* (1991) have proposed the need for linkage disequilibrium between the average degree of heterozygosity at marker loci and at quantitative trait loci (QTLs), for their association to occur.

According to Bernardo (1992) for molecular markers to be of predictive value the following conditions must be fulfilled:

1. The allele frequencies at individual loci in parental inbreds should be negatively correlated
2. dominance effects should be strong
3. trait heritability should be high
4. variation of average parental allele frequencies should necessarily be small
5. at least 30-50% of the QTLs must be linked to molecular markers
6. no more than 20-30% of the markers must be randomly dispersed or unlinked to the QTLs.

Thus, the ability to predict which combinations of inbred lines will produce the most productive hybrids is more likely to evolve from QTL analysis, which is aimed at identifying specific association between genetic loci and heterotic yield responses.

(V) MAPPING AND TAGGING OF GENES RESPONSIBLE FOR STERILITY, FERTILITY RESTORATION AND WIDE COMPATIBILITY

The improvement of even the simplest of characteristics, often requires the manipulation of a large number of genes (Flavell, 1995). Tagging of major genes is important from the practical point of view, especially for those traits that are difficult or laborious to score. A brief account of the use of molecular markers in tagging genes for wide compatibility, thermosensitive genic male sterility and fertility restoration is given below:

[A] Molecular markers for wide compatibility

The level of heterosis in hybrids depends on the genetic distance between the parents. In case of rice, adequate genetic diversity exists in indica rice varieties and, therefore, it is comparatively easier to choose good varietal combinations, which will produce significant heterosis. Japonica rice varieties, on the other hand, have very similar genetic backgrounds and, therefore, hybrids obtained from crosses between two japonica varieties do not result in significant heterosis. Generally hybrids developed from high-yielding parental lines, diverse in relatedness, ecotype and geographic origin, show promising heterosis (Yuan, 1985). However, development of hybrids by exploiting the genetic variability between japonica and indica rice varieties has resulted in increased sterility of the hybrids (Kato *et al*, 1928). Gamete abortion caused by an allelic interaction at the S-5 locus is responsible for hybrid sterility. The alleles at this locus are S-5ⁱ and S-5^j in indicas and japonicas, respectively, and some japonicas have a neutral allele, S-5ⁿ. The hybrid genotype, S-5ⁱ / S-5^j, shows gamete abortion, but the genotypes S-5ⁿ / S-5ⁱ and S-5ⁿ / S-5^j do not (Ikehashi *et al*, 1994). Thus, incorporation of the neutral allele, S-5ⁿ, known as the wide compatibility gene, into one of the parents can overcome the partial sterility commonly encountered in the progeny of indica and japonica crosses (Ikehashi and Araki, 1984; 1988). The wide compatibility gene has been incorporated into japonica types and successfully used to obtain indica-japonica hybrids (Araki *et al*, 1988; Ikehashi, 1991). In transferring this gene, however, selection in a segregating population can be carried out only after comparing the spikelet fertilities of the F1 hybrids of the individual progenies to those of the standard subspecies' testers. This work is not only time-consuming but also tedious. If these wide compatibility genes can be tagged using molecular markers, it would facilitate their transfer into elite breeding

lines and thus result in the development of fertile inter-subspecies hybrids more economically and rapidly.

The genetic mechanism of wide compatibility has been studied and the wide compatibility locus, S-5, has been located on chromosome 6, between the morphological markers, 'C' (chromogen for pigmentation) and 'wx' (Waxy endosperm) loci (Ikehashi and Araki, 1986). The chromosomal location of this wide compatibility locus has been confirmed using isozyme and RFLP markers (Liu *et al*, 1992; Zheng *et al*, 1992; Yanagihara *et al*, 1995), and the gene has been tagged via linkage to RFLP markers in a three-variety cross of wide-compatibility variety / indica // japonica (Zheng *et al*, 1992). Liu *et al*, (1992) have also used isozyme and RLFP markers to tag the wide compatibility gene and have reported that it is 9.1 map unit from RG213 and 5.9 map units from *Est-2*. Polymorphism have also been detected at the S-5 locus using the RFLP markers *Pgi-Z*, *Est-2*, RZ516, CDO475, CDO96, RG213, RG424, and RZ247. However, tight linkage at this locus has been obtained using RG213, *Est-2*, and the morphological marker, 'C'. In addition, using RG213, three distinct bands were detected, one band corresponding to each of the three respective parents in the cross, which allows unequivocal determination of the parental origin of the alleles at this locus (Brar *et al*, 1994). The S-5 locus has been determined to be approximately 1.0cM from the RFLP locus, RZ349 and 13.4cM from RZ450 (Liu *et al*, 1997). This tight linkage with RZ349 will be useful for the marker-aided transfer of this gene (chiefly the S-5ⁿ allele) to other varieties in hybrid rice breeding and for map-based cloning.

Fertility analysis of many indica-japonica hybrids has revealed a considerable variation in the fertility level of hybrids from the same wide-compatibility variety crossed to different indica or japonica genotypes (Gu *et al*, 1993; Liu *et al*, 1996). One hypothesis for the variable fertility is that additional genes are involved that modify hybrid fertility in the presence of the wide-compatibility gene, since a large number of genes exist that affect hybrid fertility in one way or another (Kinoshita, 1995). RFLP analysis has revealed the presence of two minor loci, one each on chromosomes 2 and 12, respectively, in addition to the major locus on chromosome 6 (Liu *et al*, 1997). The presence of the wide-compatibility gene on chromosome 6 alone is not sufficient to suppress hybrid sterility in indica-japonica crosses. The marker analysis of the three-way cross between the japonica wide compatibility variety, O2428, the indica variety, Nanjing 11 and the japonica variety, Balilla, has shown that heterozygosity for the indica and japonica alleles at the 2 minor

loci can collectively cause as much as 19% reduction in spikelet fertility, in the presence of the wide compatibility gene (Liu *et al*, 1997). However, the actual amount of fertility reduction maybe even greater, since the effects of these loci are certainly underestimated by marker-based analysis because of recombinations between markers and the loci for hybrid fertility. The existence of a large number of loci conferring hybrid sterility (Kinoshita, 1995; Zhang *et al*, 1997) indicates the possibility that, given the presence of the S-5ⁿ allele, different and/or additional loci may become involved in hybrid fertility in crosses of the same wide-compatibility variety with other varieties.

[B] Tagging genes for pollen sterility

Male sterility refers to the absence of functional pollen grains in otherwise hermaphrodite flowers, wherein the female gametes function normally. Male sterility is not common in natural populations and occurs sporadically, perhaps due to mutation (Singh, 1990) and is of great value in hybrid seed production. In rice, F1 pollen sterility is caused by at least six F1 pollen sterility genes. At the *S-a* locus, one of the six loci for F1 pollen sterility, the allelic interaction between *S-a_i* and *S-a_j* causes the male gametes carrying *S-a_j* allele to become abortive. Using RFLP and RAPD markers, the *S-a* locus has been mapped on chromosome 1 at distances of 6.4cM and 7.2cM from RFLP markers, CDO548 and RG146, respectively and 6.8cM and 11.2cM from RAPD markers, O11-1000 and Y13-500, respectively (Zhuang *et al*, 1999). Mapping of the *S-a* locus is an important step towards marker-aided selection for overcoming hybrid sterility in rice.

[C] Fertility restoration using the nuclear restorer gene

The nuclear restorer gene (*Rf*) can counteract the pollen sterility and restore fertility in the cytoplasmic male sterile (CMS) lines (Newton, 1988). The fertility restorer gene is a dominant gene, which is indispensable for hybrid seed production systems, because F1 hybrid CMS plants are fertile when they have heterozygous *Rf* gene. The fertility restoring ability varies according to the number and strength of the dominant restorer genes and the nature of the modifier complex, which varies with the background of the CMS lines. Certain restorers are governed by a single dominant gene, while others have two dominant genes (Siddiq *et al*, 1994). The penetrance and expressivity of the fertility restorer gene depend mainly on the environment and nature of the CMS lines used (Siddiq *et al*, 1994).

Cytoplasmic male sterility also occurs in rice, when the cytoplasm of japonica rice is replaced with that of indica rice or wild rice (Katsuo and Mizushima, 1958; Li and Zhao, 1986; Virmani and Shinjyo, 1988). The CMS system that occurs in combination between the cytoplasm of an indica rice, Chinsurah Boro II and the nucleus of a japonica rice, Taichung 65, is called the ms-bo-type or BT-type. The BT-type CMS system is restored by the nuclear restorer gene, *Rf-1*, which has initially been identified in Chinsurah Boro II and mapped on chromosome 10 (Shinjyo, 1975; 1984). It has been proposed that *Rf-1* is involved in the efficiency of processing or editing mitochondrial transcripts in rice (Kadowaki *et al*, 1990; Iwabuchi *et al*, 1993; Akagi *et al*, 1994).

Development of new restorer strains by conventional breeding requires repeated back-crossing of the restorer lines with the recurrent parent, followed by selection of the fertile plants, which is an extremely laborious and time consuming process. If molecular markers could be employed to tag the restorer genes, it would reduce the time required to develop new restorer lines.

The nuclear restorer gene for fertility in rice, *Rf-1*, essential for hybrid seed production, has been tagged using ISSR marker (AG)8YC (UBC 835). Two bands of 420bp and 400bp have been obtained in the two near-isogenic lines, with or without, the *Rf-1* gene, respectively. Cloning followed by sequencing of these fragments has revealed that, tetranucleotide repeats, (GATG)_n, are responsible for generating polymorphism at this locus. Based on the sequencing data, a new set of primers, which specifically amplify this inter-repeat region have been synthesized, and the locus amplified by this new primer set has been named, OSRRf (*Oryza* simple sequence repeat, *Rf-1*). The OSRRf marker is tightly linked to the *Rf-1* gene at a distance of 3.7[±] 1.1cM (Akagi *et al*, 1996). This codominant marker will not only be useful for breeding restorer lines but will also accelerate the breeding of maintainer lines by eliminating the *Rf-1* gene. It can also be applied to seed purity management of hybrid rice seeds by eliminating contamination of CMS lines, maintainer lines and restorer lines from F1 hybrid rice at seedling stage.

[D] Identification of genes for environment-sensitive genic male sterility

The cytoplasmic male sterility (CMS) system involves the use of three lines namely, the cytoplasmic male sterile (CMS) line, the maintainer line and the restorer line is most effective for developing rice hybrids. However, this system is very cumbersome as it involves the use of three different lines and is

restricted only to those germplasms where maintainer and restorer lines are available abundantly. Continuous use of a CMS system may result in the hybrids becoming potentially vulnerable to a biological stress. Also the availability of only a few stable CMS lines with good combining ability and the negative effects associated with sterile cytoplasm are some of the major limitations of the three-line system. In rice, a novel genic male sterility, called photoperiod-sensitive genic male sterility (PGMS), has been reported which can revert back to fertility under certain photoperiods (Shi, 1981; 1985; Shi and Deng, 1986). Later on thermo-sensitive genic male sterility (TGMS) has been discovered, which can revert to partial or total fertility under certain temperature conditions (Zhou *et al*, 1988; 1991; Muruyama *et al*, 1990a; 1990b; Wu *et al*, 1991). Based on these two systems, PGMS and TGMS, a new strategy not involving the maintainer line has been put forth, in which any fertile line can be used as pollen parent to develop a rice hybrid (Yuan, 1987). This method of hybrid rice breeding is called the two-line system. The use of TGMS and PGMS systems in two-line breeding is simple, inexpensive, efficient and eliminates the limitations associated with the CMS system. Thus, the discovery and application of both TGMS and PGMS systems have a great potential for revolutionizing hybrid seed production.

(D.1.) Photoperiod-sensitive genic male sterility (PGMS)

PGMS lines show complete pollen sterility in long-day conditions under certain temperature regimes, but revert to fertility under short day conditions. The critical day-length for PGMS lines is 13.75-14.00 h/d and when the day-length is 14.00 h/d or longer, these lines show sterility (Zhang, *et al* 1987). Twilight has the same effect as day-length, on inducing fertility alteration in these lines (Zhang and Yuan, 1989). Temperature change during the day can also modify the fertility alternation in these lines (He *et al*, 1987). Thus, both day-length and temperature are responsible for regulating fertility alterations (Zhang *et al*, 1992a), and the effects of photoperiod and temperature on fertility alteration are interdependent (Zhang *et al*, 1992b). Under short day conditions, the critical day length for fertility decreases as the temperature rises, and it becomes longer with higher fertility as the temperature falls. For hybrid seed production, the ideal sterile line should have a low critical temperature point for sterility induction, a wide range of photoperiod sensitivity and a strong supplementary effect between photoperiod and temperature (Yuan, 1992b; Zhang and Yuan, 1992). If the genes controlling photoperiod sensitivity in such an ideal line can be tagged, it would help to

transfer these genes by marker-assisted selection. In rice, two PGMS genes designated *pms-1*, and *pms-2* located on chromosomes 7 and 2, respectively, have been tagged (Zhang *et al*, 1994b). In tropical areas, where the day-length differences are small but temperature differences exist between low and high altitudes, TGMS system is considered more useful than the PGMS system (Virmani, 1992).

(D.2.) Thermosensitive genic male sterility (TGMS)

In TGMS lines, pollen fertility alterations are regulated by temperature, wherein high temperature induces sterility while low temperature induces fertility or semi-sterility (Sun *et al*, 1989; Wu *et al*, 1991; Yang *et al*, 1989; Zhou *et al*, 1991; Virmani, 1992; Maruyama, *et al*, 1992). Reverse TGMS lines on the other hand are sterile at lower temperatures and fertile at higher temperatures. The critical temperatures determining fertility / sterility vary greatly depending on the source of the gene. Tagging of the TGMS genes using molecular markers, will aid in transferring these genes to other lines and thus help to accelerate hybrid seed production.

Wang *et al* (1995b) have been the first to report on mapping of the rice TGMS gene. They have used bulked segregant analysis of the F₂ population and identified five polymorphic RAPD products, of which only two fragments amplified by OPT02 and OPB-19 have single copy sequences. These two fragments of 0.5 Kb (OPT02) and 1.2Kb (OPB19) have been cloned and mapped at distances of 12.9cM and 12.6cM, respectively, on either side of RZ562. The TGMS gene has been located on chromosome 8 due to the presence of RZ562 linked to TGMS gene on this chromosome (Wang *et al*, 1995a). AFLP markers have been used to develop more markers linked to the TGMS gene, in order to develop a detailed physical map of the gene-encompassing region. Six polymorphic AFLP products have been cloned and sequenced, and used to screen rice BAC libraries (Wang *et al*, 1998). Lang *et al* (1999) have used RAPD to design specific primers for PCR amplification. Using the RAPD marker, OPF18₂₆₀₀, two specific primer combinations from different positions of the sequence have resulted in generation of codominant STS markers, which have been found to be tightly linked to the *tms3(t)* gene at a distance of 2.7cM. The marker-assisted selection of this trait has been estimated at 84.6%, indicating that the PCR-based markers for *tms3(t)* can be used to select TGMS plants at the seedling stage in the segregating populations

in environment independent of controlled temperature regime (Lang *et al*, 1999).

Using the F₂ population of a cross between a TGMS indica mutant, TGMS-VN1, and fertile indica line, CH-1, a new TGMS gene, tentatively designated *tms4(t)* has been identified (Dong *et al*, 2000). Four AFLP markers (E2/M5-600, E3/M16-400, E5/M12-600 and E5/M12-200) and one STMS marker, RM227 have been found to be linked to the TGMS gene (Dong *et al*, 2000). The AFLP marker, E5/M12-600, is linked to the TGMS gene at a distance of 3.3cM on chromosome 2 (Dong *et al*, 2000). PCR using primers flanking the E5/M12-600 marker will help in marker assisted transfer of this gene to different genetic backgrounds. Reddy *et al*, (2000) have used AFLP, RAPD, and STMS markers to characterize diverse TGMS lines. The STMS marker RM257 and a 330bp fragment obtained on AFLP, with the combination EAA/MCAG, closely segregate with the TGMS gene at distances of 6.2cM and 5.3cM, respectively, on chromosome 9. The 330bp fragment obtained by AFLP, when converted into STS marker, TS200, has produced a dominant polymorphism specific to TGMS. The STMS marker, RM257 has produced a codominant polymorphism with 145bp and 132bp products in sterile and fertile lines Mapping and tagging of genes responsible for sterility, fertility restoration and wide compatibility, respectively (Reddy *et al*, 2000). Thus, both the marker TS200 and RM257 located on either side of the TGMS gene locus, are very useful for marker-assisted selection.

(VI) QUANTITATIVE TRAIT LOCI FOR HETEROSIS / YIELD

Most of the agronomically and economically important traits like yield and yield components (grain number, grain weight, plant height, days to flowering, maturity date, *etc.*) are controlled by a relatively large number of loci, each of which makes a small positive or negative contribution to the final phenotypic value of the trait. Such loci are called as quantitative trait loci (QTLs), and the traits showing continuous variation in phenotype are termed as polygenic traits, because the final expression of the phenotype is governed by the genetic variation at a large number of loci, modified by environmental effects. While the theory and techniques of quantitative genetics (Falconer, 1960) have proven useful in the study of quantitative traits, these characters continue to be more difficult to manipulate than single gene traits in breeding programs (Tanksley *et al*, 1989). However, if these complex traits can be resolved into their individual genetic components, it might be possible to deal

with these characters with the efficacy of single gene traits (Tanksley *et al*, 1989). The progress in DNA marker technology followed by the subsequent development of molecular linkage maps has made it possible to identify, map and measure the effects of genes underlying quantitative traits (Tanksley, 1993; Dudley, 1993). By using DNA markers for QTL analysis, it is possible to distinguish individual genetic components that are sometimes masked by the interaction of major genes and by the environment (Ikeda and Wasaka, 1997). With molecular markers, it is possible to assign chromosomal positions to individual QTLs, to determine the types and magnitude of gene effects of individual QTLs, and also to determine which parent possesses the positive allele at each QTL (Edwards *et al*, 1987; Tanksley *et al*, 1989; Lander and Botstein, 1989; Paterson, 1995; Stuber, 1995). Also, DNA markers linked to QTLs are exceedingly valuable for genetic diagnostics, as the breeding processes in quantitative traits require an enormous time frame and logistics. The use of DNA markers obviates most of these difficulties.

The ability to detect a QTL with a molecular marker is the function of the magnitude of the QTLs' effect on the character, the size of the population being studied and the recombination frequency between the marker and the QTL (Tanksley *et al*, 1989). If the marker and the QTL are located far apart, then the possibility that they will be transmitted together to the progeny individuals will be reduced due to recombination events. Using markers, each QTL can be studied as a discreet entity, and its individual and interactive properties can be measured. Molecular markers also permit the precise and rapid transfer of QTLs into superior crop varieties or genetic stocks. Some of the important QTLs have been discussed below.

(A) Grain yield

The performance of any cross is normally estimated on the basis of the final grain yield, which is in turn affected by several other yield contributing traits. Grain yield is a function of three yield components, number of plants per unit area, number of grains per plant and grain weight. Identification of QTLs with a major effect on heterosis for grain yield will facilitate the development of highly productive hybrid varieties.

Stuber *et al* (1992) have mapped QTLs associated with seven major traits, including grain yield, in a cross between two widely used elite maize inbred lines B73 and Mo17. The analysis has been carried out to explore heterosis or hybrid vigor and genotype-by-environment interaction, which are

two important phenomena in maize genetics. Seventy six markers, representing 90-95% of the maize genome have been analyzed using single-marker method and interval mapping method. Both methods have shown virtually the same results in detecting QTLs affecting grain yield throughout the genome, except on chromosome 6 (Stuber *et al*, 1992). For the other quantitative traits, fewer QTLs have been detected. This study has demonstrated that heterozygous, rather than homozygous, marker loci are often most closely associated with the highest yielding progeny. Thus, overdominance (or pseudoverdominance) plays an important role in heterosis and QTL analysis is capable of identifying specific regions of the genome that contribute to heterotic responses.

In rice, Xiao *et al* (1995) have analyzed QTLs for yield-related traits by single point analysis and interval mapping using RFLP markers. A total of 37 significant QTLs ($LOD \geq 2$) have been detected with 141 RFLP markers on backcrossing a set of F7 lines from a subspecific cross showing strong F1 heterosis to the two parents. 73% QTLs have been detected in one of the BC1 F7 populations. These include QTLs for plant height, days to heading, days to maturity, panicle length, panicles per plant, spikelets per panicle, grains per panicle, percent seed set, 1000 grain weight, spikelets per plant, grains per plant and grain yield. Only two significant QTLs have been detected for grain yield, which may be because of severe spikelet infertility (as measured in percent seed set) that has been observed in both the backcross populations. In 82% of these cases the heterozygotes are found to be superior to the respective homozygotes. The remaining 10 QTLs (27%) have been detected in both the BC1F7 populations, with the heterozygote having a phenotype falling between those of the two homozygotes, and in no instances have the heterozygotes been found to be superior to the homozygotes. This suggests that dominance complementation is the major basis of heterosis in rice. This correlation has also been strengthened by the fact that no correlation has been observed between most of the traits and overall genome heterozygosity, and that there are some recombinant inbred lines in the F8 population having phenotypic values superior to the F1 for all the traits evaluated, which is not expected if overdominance is a major contributor of heterosis (Xiao *et al*, 1995).

Li *et al* (1997) have identified 19 QTLs for yield traits, using RFLP markers, in the F4 progeny of a subspecific cross. Of these, eight QTLs for kernel weight, six for grain number and two for grain weight per panicle have been identified. Of the 63 markers distributed throughout the genome, that

appear to be involved in the 73 highly significant interactions, 46 (73%), do not appear to have "main" effects on the relevant traits but influence the trait(s) predominantly through interactions. Li *et al* (1997) have, therefore, suggested that epistasis is the important genetic basis for complex trait such as yield, especially of low heritability components like grain number and grain weight per panicle. Recently, Li *et al* (2000b) have constructed a genetic linkage map comprising 151 molecular markers for QTL mapping from a population derived by crossing Zenshan 97 and Minghui 63, the parents of Shanyou 63, an elite rice hybrid widely grown in China and have detected a total of 20 distinct QTLs.

Although F₂s are the most informative population for genetic analysis, it has been difficult to use F₂ population directly for QTL analysis because of the difficulty in assessing the reliability of the data, due to an inability to estimate the experimental errors. Li *et al*, (2000b) have, performed QTL analysis for yield and yield-component traits of an F₂ population of rice based on data from replicated field trials over two years, using vegetative shoots of ratooned plants, making use of the ratooning habit of rice. They have demonstrated that vegetatively ratooned F₂ populations may have a considerable utility in the mapping of QTLs, especially for dominant types of gene actions. Most of the QTLs detected in the ratooned F₂ population have also been detected in the F₂:3 population.

[B] QTLs for seedling vigor

Seedling vigor, which is the ability of a plants' aerial parts to emerge rapidly from soil or water (Heydecker, 1960), is a desirable trait to incorporate into modern rice cultivars. Retardation of seedling growth at low temperature is a problem commonly occurring in temperate rice growing areas and in tropical and sub-tropical areas at high elevations or with cold irrigation water-supply. In such regions, a delayed emergence of the rice seedlings from water greatly increases seedling mortality (Peterson *et al*, 1978). Vigorous cultivars are also needed for large rainfed and upland areas in the tropics where dry-seeding is practiced and competition from weeds is a serious problem. Also, the increasing importance of direct seeding in many Asian countries (Dingkuhn *et al*, 1992) has made it essential to improve seedling vigor of rice cultivars. In parts of southern USA, where rice is drill-seeded good seedling vigor is important for adequate stand establishment. Several quantitative traits have

been associated with seedling vigor in rice. In water-seeded rice cultures, long shoots have been associated with seedling vigor (Peterson *et al*, 1978) while rapid root growth is important for proper soil anchorage to reduce seedling floatation, (Williams and Peterson, 1973). Long mesocotyls and coleoptiles promote seedling emergence in drill-seeded rice culture (Turner *et al*, 1982; Dilday *et al*, 1990). In spite of the genetic variation for seedling vigor in rice (Jones and Peterson, 1976; Mgonja *et al*, 1993), breeders have had difficulty in improving the seedling vigor in the semidwarf cultivars (McKenzie *et al*, 1994). The poor success achieved by breeders using conventional breeding methods, in both temperate (McKenzie *et al*, 1994) and tropical rice growing areas (Herdt, 1991) could be partly due to the traits' association with undesirable characters such as, tallness, lodging susceptibility, large grain size and earliness (Peterson, *et al*, 1978; Li and Rutger, 1980), that are selected against during the breeding process. If the QTLs underlying seedling vigor can be studied using molecular markers, it will help to elucidate the genetic basis of this trait and, therefore, improve the breeding and selection efficiency.

An F₂ population of a cross between a low-vigor japonica cultivar, Labelle (LBC) and a high vigor indica cultivar, Black Gora (BG) has been screened with RFLP markers and a genetic linkage map of 117 markers has been constructed (Redona and Mackill, 1996a). Using interval analysis, 13 QTLs have been identified, each accounting for 7-38% of the phenotypic variance. Of these, four QTLs are responsible for controlling shoot length, 2 each control root and coleoptile lengths and five influence mesocotyl length. Seven of the twelve rice chromosomes contain at least one of the 13 QTLs and certain regions of the genome influence more than one trait. For example, the QTLs for length of shoot, coleoptile and mesocotyl, map to the same or adjacent intervals on chromosome 3, while QTLs for root and shoot length map to adjacent intervals on chromosome 1. This can be either due to linkage or pleiotropy, but linkage is a more plausible explanation when adjacent QTLs have alleles from different parents, as seen on chromosome 3 (Redona and Mackill, 1996a). Two of the QTLs identified in the above study, viz. the mesocotyl length QTL on chromosome 1, and the shoot-length QTL on chromosome 9, map to locations close to members of the alpha-amylase gene family, which have been implicated in rice breeding vigor (Ranjhan *et al*, 1991; Causse *et al* 1994; Thomas and Rodriguez, 1994). On the basis of interval analysis, Redona and Mackill (1996a), have observed that the rice cultivar, BG has no positive alleles for shoot length, which is the best vigor

indicator in water-seeded rice (Jones and Peterson, 1976), indicating that the trait is controlled by many loci, most of which have minor effects. This is supported by the identification of additional alleles, one each on chromosomes 1, 3, and 6, for shoot length (Redona and Mackill, 1996a). These results suggest that the lack of QTLs with major effects for seedling-vigor related traits, coupled with the loose relationship between parental genotype and phenotype, makes the identification of superior donors and breeding lines based only on phenotype extremely difficult. QTL analysis to identify superior donors and marker-aided selection strategies may be useful in breeding for rice seedling vigor.

Redona and Mackill (1996b) have used 129 RAPD and 18 RFLP markers on the F₂ population of a cross between two japonica cultivars, Italica Livorno (IL), with high seedling vigor and Labelle (LBL), with low seedling vigor to detect QTLs for seedling vigor. Two coleoptile, five root and five mesocotyl length QTLs, each accounting for 9-50% of the observed phenotypic variation have been identified by interval analysis. About two thirds of the alleles positive for the putative QTLs are obtained from the high vigor parent, IL. One RAPD marker, OPAD13-720 is associated with the IL allele that accounts for 18.5% of the phenotypic variation for shoot length, the most important determinant of seedling vigor in water-seeded rice. These results indicate that RAPDs are useful for map development and QTL mapping in rice populations with narrow genetic base, such as those derived from crosses among japonica cultivars (Redona and Mackill, 1996b). However, further identification and tagging of QTLs from other sources for seedling vigor related characters might be necessary if genetic gain from selection in breeding for these quantitative traits is to be maximized.

[C] Plant height

Plant height is an important character associated with productivity of rice where more than 50 major genes for dwarfism or semidwarfism have been identified in rice (Kinoshita and Takahashi, 1991). The semidwarf gene, *sd-1*, is the best characterized of these and has been extensively used to produce high-yielding semidwarf varieties (Hargrove *et al*, 1988; Rutger and Bollich, 1989). The other semidwarf genes have not been utilized because they have been associated with poor agronomic performance (Rutger, 1984; Kikuchi and Ikehashi, 1984).

Quantitative variation in plant height has frequently been observed in plant breeding populations, suggesting that there must be a number of other loci, which affect the height of the rice plants. This has been observed in the rice variety Maybelle, which is almost as short as the semidwarf varieties, but yet doesn't contain any known semidwarf genes (Bollich *et al*, 1991). Austin and Lee (1994) have analyzed the QTL for plant height on the long arm of chromosome 1 of maize, using RFLP markers. They have observed that, in the F3 generation, the QTL comprises one large genome region, while in the F6 generation, the region is resolved into three distinct QTLs, defined by the same RFLP loci as in the F3 generation. The additional opportunities for recombination and the increased homozygosity of the F6 lines have facilitated resolution of the region.

In rice, 4 QTLs which alter plant height from 4-7cm and have mapped to 4 of the 12 rice chromosomes, have collectively explained 48.8% of the observed phenotypic variation. The 2 QTLs with largest effects, QPh3a and QPh8a, have individually explained 21% and 25% of the genotypic variation, respectively, while the two minor QTLs, QPh2a and QPh9a, have explained 7.9% and 8.4% of the genotypic variation, respectively (Li *et al*, 1995). None of the QTLs for plant height have been mapped to chromosome 1, the location of the semidwarf gene, *sd-1* (Khush and Kinoshita, 1991). The QTLs for plant height are located near RG418a on chromosome 3, and on chromosome 12, between RG323 and RG101 and between RG241a and RZ397. Three of the QTLs for plant height, QPh2a, QPh3a and QPh8a, have partial dominance towards shortness and the fourth QTL, QPh9a, exhibits overdominance towards plant height. The observed overdominance at the fourth QTL for plant height, may be partly responsible for the observed heterosis in plant height (Li *et al*, 1995).

Xiao *et al* (1995) have detected 5 QTLs for plant height in rice, of which three QTLs on chromosomes 2, 3, and 8 have resulted in increased plant height for the heterozygote, as compared to the homozygote, while 2 QTLs on chromosomes 6 and 7, have caused decreased plant height for the heterozygote. The QTL on chromosome 3 bordered by RFLP markers, XNpb249 and RZ16, is known to account for 26% of the total phenotypic variation. In a second population, 3 QTLs have been detected for plant height at approximately the same positions on chromosomes 5, 6 and 8. However, the heterozygotes have shown opposite effect from that observed in the first population for the respective QTLs, suggestive of additive gene action.

Several QTLs that have a significant impact on plant height have been detected in maize by using molecular markers (Helentjaris, 1988). In an effort to identify QTLs for plant height, one of the loci has been found near the centromere of chromosome 9, which could account individually for at least 27% of the variance of this trait (Helentjaris and Shattuck-Eidens, 1987).

[D] Heading date

Li *et al.*, (1995) have identified the QTLs for heading date using RFLP markers on the F4 progeny of a cross between Lemont and Teqing, both of which differ in the heading date by only 6 days. Two of the QTLs, QHd3a, and QHd8a located on chromosomes 3 and 8, respectively, have explained greater than 40% of the total genotypic variation in heading date. The QTLs for heading date have mapped to approximately the same genomic locations as QTLs for plant height. (Li *et al.*, 1995). The chromosomal locations of the 3 QTLs for heading date, do not correspond to those reported for the mapped maturity genes *Ef-1*, *Se-1*, and *Se-2* (Kinoshita and Takahashi, 1991). Thus, the observed heterosis in the F1 plants for heading date can be largely explained by the dominance for earliness of the identified heading date loci and the distribution of earlier heading alleles in the parents (Li *et al.*, 1995). Xiao *et al.* (1995) have tagged QTLs for heading date in rice using RFLP markers by backcrossing a set of 194 F7 lines derived from a subspecific rice cross showing strong F1 heterosis in the two parents. The heterozygotes have shown reduced days to heading for the 2 QTLs on chromosomes 3 and 4 in one population, and increased days to heading for two QTLs on chromosomes 3 and 7 in another population.

Thus, in conclusion, several QTLs for heterosis have been identified and detailed information has been gathered about the loci involved in the different yield components. This knowledge will be very useful to improve yield potential by manipulation of these loci through marker assisted selection. The molecular markers associated with QTLs will also be useful in various stages of varietal development to screen parental material and breeding lines for desired genotypes at marker loci in order to introgress and/or pyramid favorable QTL alleles into new populations.

[E] QTLs from wild germplasm

Although wild and unadapted germplasm is phenotypically less desirable than modern varieties in its overall appearance and performance, it is

a unique source of genetic variation and can be used for the improvement of simply inherited traits such as disease and insect resistance, cytoplasmic male sterility etc. *O.spontanea* has been used as a source of wild abortive CMS, which has provided the basis of modern hybrid rice (Li and Zhu, 1988). However, in wild germplasm the superior trait of interest cannot be identified phenotypically making it virtually impossible to utilize wild accessions for improvement of quantitatively inherited traits. For most quantitative traits, phenotype is conditioned by several genes, having positive trait enhancing or negative trait depressing alleles. In elite cultivars, positive alleles are present in high frequency, while in wild germplasm these desirable alleles are present in low frequency and are often masked by the effects of other deleterious alleles.

Tanksley and Nelson (1996) have used a strategy called advanced backcross QTL analysis to simultaneously discover and transfer valuable QTLs from unadapted germplasm into elite breeding lines. Using the same method, Tanksley *et al* (1996) have exploited the potential of a wild species, *L.pimpinellifolium*, for improving yield and quality of an elite breeding line of tomato.

In order to identify trait improving QTL alleles from wild, exotic germplasm, an accession of *O.rufipogon*, a relative of cultivated rice, has been used to develop an interspecific BC2 testcross population, which has then been evaluated for 12 agronomically important quantitative traits using 122 RFLP and microsatellite markers (Xiao *et al*, 1998). The *O.rufipogon* accession has been observed to be phenotypically inferior for all 12 traits. However, transgressive segregants that outperformed the original elite hybrid rice variety, have been obtained for all the traits examined and have been used to identify QTLs. Of the 68 significant QTLs identified totally, 35 (51%) have beneficial alleles derived from the phenotypically inferior *O.rufipogon* parent, of which 19 (54%) are free of deleterious effects on other characters. Xiao *et al* (1998) have identified eight QTLs each, for 1000 grain weight and days to maturity, seven QTLs each controlling days to heading, panicle length, percentage seed set and grain yield, six QTLs each for plant height and grains per plant, five QTLs influencing grains per panicle, four QTLs influencing spikelets per panicle, two QTLs controlling panicles per plant, and lastly one QTL influencing spikelets per plant. *O.rufipogon* alleles at the two QTLs on chromosomes 1 and 2 have been associated with 18 and 17% increase in grain

yield per plant, respectively, without delaying maturity or increasing plant height.

The results indicate that, despite their overall inferior appearance, wild species contain QTL alleles that are likely to substantially improve agronomically important traits, including yield. This implies that the wild and unadapted germplasm reservoir can hold the key to future productivity increases in rice and other crop species.

(VII) MARKER-ASSISTED SELECTION FOR HETEROSIS

Marker-assisted selection (MAS) is based on the concept that it is possible to infer the presence of a particular gene/QTL from the presence of a marker tightly linked to the desired gene/QTL. If a DNA marker linked to the target gene can be identified, it will be useful to infer the presence of the phenotypic trait by screening for linked DNA markers in all individual plants of a breeding population. This indirect selection of target traits by DNA markers could be very effective in backcross breeding, where undesirable chromosomal segments (ie most of the donor chromosome segments except for the gene/QTL to be introgressed) must be removed within several backcross generations (Ikeda and Wasaka, 1997). Using MAS, the breeder can test individual plants after extracting small amount of DNA from each of the plants without destroying the original plant. Presence or absence of the marker band in the agarose gel will determine the genotype of the individual plants. Only the plants, which are in later, advanced generations, will need to be tested for the phenotypic traits of interest. Thus, using MAS, the breeder can conduct many rounds of selection in one year, without depending on the natural occurrence of the gene/QTL.

The predictive value of genetic markers used for MAS depends on their inherent reproducibility (Weeden *et al*, 1992), map position, linkage with economically important traits (qualitative or quantitative) and their ability to cosegregate or be closely linked with the desired trait. Similarly, selection of multiple loci or QTLs, using molecular markers, can be effective if a significant association is detected between the quantitative trait and markers (Edwards *et al*, 1987; Lande and Thompson, 1990; Edwards and Page, 1994).

Stuber and Edwards (1986) have compared MAS using 15 isozyme loci in two F₂ maize populations with phenotypic selection and have observed that gain from MAS is equal to that from phenotypic selection. Stuber *et al*

(1992) predicted that MAS could be effective for transfer of desired factors at multiple loci for traits such as grain yield, if the entire genome is covered instead of only 30-40% studied by Stuber and Edwards (1986). According to Johnson (1991) MAS of S4 lines, using marker associations based on S2 test performance for grain yield, is more predictable between years and testers than selection based on S2 testcross performance alone. Stromberg (1992) has compared selection based on F2 testcross mean yields with MAS among F2 plants, followed by selection among S4 plants within lines derived from the selected F2 plants. MAS of semidwarfing gene, *Sd-1*, has been carried out in rice (Cho *et al*, 1994). A prerequisite to using DNA markers, is saturation of regions encompassing the locus of interest, in the genetic linkage map (Edwards *et al*, 1987). Gu *et al* (1995) have used the codominant marker, alcohol dehydrogenase (*Adh-1*) gene in pea and applied it for screening photoperiod genes in common bean by MAS without the need for electrophoresis.

Thus, the integration of molecular marker technology into existing plant breeding programs the world over will allow researchers to access, transfer and combine genes more rapidly and precisely than before. In case of QTLs, the success of MAS depends on the ability of the markers to detect QTLs and the consistency of QTLs over environments and generations (Lande and Thompson, 1990; Shoemaker *et al*, 1994). Detailed information about the loci involved in the different yield components will be helpful in improving yield potential by deliberately manipulating these loci through MAS. Thus, in conclusion, MAS is a potential strategy of increasing selection efficiency by allowing for earlier selection (from seedling stage of the plant) and reducing plant population size used during selection.

(VIII) FUTURE TRENDS FOR INCREASING HETEROSIS

[A] Testing purity of hybrid seeds

Seed purity is closely related to their heterosis, performance and yield increase. Therefore, for seed commercialization of hybrid crops, examination of the purity of seeds is essential before field-production. In rice, DNAs of Gangyou-22, a major hybrid rice in China, and its parents Gang46A (CMS line) and CDR22 (restorer line), have been screened with microsatellite primers to determine polymorphic markers. The STMS marker, RM168 has produced polymorphic bands specific to each of the two parents but different from the other 22 restorer lines and nine cultivars, indicating that it can be accurately

and efficiently used to examine the purity of a hybrid rice variety at an early stage (Li *et al*, 2000a). Sonti *et al* (2000) have also tested the purity of rice hybrids using DNA markers.

Such studies can be extended to the rice hybrids from other parts of the world and to hybrid varieties of other crops also. If DNA can be isolated from seeds instead of seedlings it will further simplify the procedure and eliminate the need for growing seeds upto seedling stage and thus prevent loss of resources. Such rapid, nondestructive DNA isolation methods half seeds by Wang *et al* (1993) and Zhai *et al* (1996) have made it easier to implement MAS. Also, if two markers can be multiplexed together in one single polymerase chain reaction, it will further improve the accuracy of seed purity examination.

[B] Pyramiding of genes and/or QTLs

Considerable research effort has been directed to identifying genes/QTLs for heterosis and in transferring them into new cultivars and hybrids. However, if several genes and/or QTLs from genetically diverse sources could be introduced into a single cultivar or hybrid, it may result in a line with very high heterosis. Several genes conferring resistance to bacterial blight of rice, when incorporated into a single cultivar, have resulted in a more durable, broader spectrum and higher level of resistance than lines containing only a single resistance gene (Yoshimura *et al*, 1995; Huang *et al*, 1997). Similarly, pyramiding of genes/QTLs for yield/heterosis, could be a useful strategy for generating hybrids with a higher yield potential and better heterosis. However, it may be difficult or even impossible to select plants containing multiple genes/QTLs based on phenotype alone, because the action of one gene/QTL may mask the action of another. Marker-assisted selection is of enormous use in gene/QTL pyramiding where the presence of one gene/QTL has to be confirmed. Thus, using DNA markers, it is possible to unambiguously select pyramid lines.

[C] Map-based cloning of genes

Map based cloning or positional cloning, also known as reverse genetics, is a strategy which requires knowledge about the chromosomal location of the gene but does not require prior knowledge about the gene or its products (Young, 1990; Wicking and Williamson, 1991; Collins, 1992). Map

based cloning and transposon tagging are methods which are being employed to isolate genes corresponding to desirable traits of agronomic importance. The first step in map based cloning is the identification of markers tightly linked to the gene and on either side of the gene. A combination of genetic analysis and molecular tools will then facilitate the isolation of genes corresponding to the trait of interest. Map based cloning has been successfully used to isolate the *Pto* gene in tomato (Martin *et al*, 1993) and the *Xa21* and *Xa1* genes in rice (Song *et al*, 1995, Yoshimura *et al*, 1996).

If genes related to heterosis can be cloned by map-based cloning, it would help to elucidate some of the biological complexities of important phenotypes and create unforeseen opportunities for their manipulation and utilization in plant breeding strategies. If these new genes, isolated by positional cloning can be utilized in breeding programs, it will result in genetic gain, increased heterosis, and thus, increased agricultural productivity. Helentjaris and Heun (1994) have even suggested the cloning of QTLs and the use of such cloned QTLs in genetic engineering strategies for crop improvement.

[D] Comparative mapping/syntenic mapping of genes/QTLs for heterosis

Considerable synteny has been observed in the genomes of several crop plants, such as tomato, potato and pepper (Tanksley *et al*, 1988; 1992); wheat, rice and maize (Ahn *et al*, 1993), wheat, barley and rye (Devos *et al*, 1993a), sorghum and maize (Pereira *et al*, 1994) and *Arabidopsis* and *Brassica* (Teutonico and Osborn, 1994). Comparative genetic studies using rice, wheat, maize, oat, sorghum, foxtail millet and sugarcane have demonstrated that gene content and gene order are highly conserved between species within the grass family, both at the map and megabase level (Devos and Gale, 1997).

Fatokun *et al* (1990), have shown that the regions controlling seed weight genes in mungbean and cowpea are conserved and the order of loci in these blocks have, by and large, remained unchanged. The loci controlling heading date in hexaploid oat (chromosome 5) and rice (chromosome-3) are homologous (Causse *et al*, 1994) while those controlling heading date in hexaploid oat (chromosome 17), vernalization gene *Vrn5* in wheat (chromosome 7A) and the photoperiod response gene, *Se1* in rice (chromosome 6) are homologous. Comparative mapping in maize and sorghum has revealed three putatively orthologous regions for plant height (Pierera and Lee, 1995). Syntenic relationships can help to predict the positions of the orthologous genes

of agronomic importance in related species (Jena, *et al*, 1994; Harrington *et al*, 1997). Identification of heterologous DNA markers tightly linked to the gene of interest will be useful to monitor introgression of chromosomal segments in breeding programs. If related species show conserved linkages in regions of shared orthologous loci, such genes can be characterized more efficiently and cloned in those species that have well-developed genetic linkage maps and small and less complex genomes.

Once the gene in the source species has been cloned and sequenced, this information can be used to quickly isolate the orthologous gene in the related species having larger genome size. Thus, map information about the gene-rich species may provide valuable information about a species and vice-versa. Comparison of locus order and distribution of recombination events may be useful to suggest strategies to incorporate this germplasm in wide crosses (Devos *et al*, 1993b) which will help to considerably improve the gene pool of the crop.

In summary, the phenomenon of heterosis is extremely complex and although molecular markers have added to our understanding of this phenomenon, further studies need to be performed before we can draw any conclusions. The completion of rice and Arabidopsis genome sequencing as well as the current progress in sequencing other crop species will help to develop newer and better markers for MAS of heterosis.

ACKNOWLEDGEMENTS

APD acknowledges the Council of Scientific and Industrial Research, New Delhi, India for a Senior Research Fellowship.

REFERENCES

- Ahn S, Anderson JA, Sorrells ME, Tanksley SD (1993) Homeologous relationships of rice, wheat and maize chromosomes. *Mol Gen Genet* 241: 483-490
- Ajmoné Marsan P, Castiglioni P, Fusari F, Kuiper M, Motto M (1998) Genetic diversity and its relationship to hybrid performance in maize as revealed by RFLP and AFLP markers. *Theor Appl Genet* 96: 219-227
- Akagi H, Sakamoto M, Shinjo C, Shimada H, Fujimura T (1994) A unique sequence located downstream from the rice mitochondrial *atp σ* may cause male sterility. *Curr Genet* 25: 52-58
- Akagi H, Yohozeki Y, Inagaki A, Nakamura A, Fujimura T (1996) A co-dominant DNA marker closely linked to the rice nuclear restorer gene, *Rf-1*, identified with inter-SSR fingerprinting. *Genome* 39: 1205-1209

- Akagi H, Yokozeki Y, Inagaki A, Fujimura T (1996) Microsatellite DNA markers for rice chromosomes. *Theor Appl Genet* 93: 1071-1077
- Araki H, Toya K, and Ikehashi H (1988) Role of wide compatibility genes in hybrid rice breeding. In: *Hybrid Rice*. International Rice Research Institute, Manila, Philippines. pp 79-83
- Austin DF, Lee M (1994) Comparative linkage analysis of RFLP loci and QTL in F2:3 and F6:7 recombinant inbreds. *Maize Genet Coop Newsl* 68: 7-8
- Barbosa-Neto JF, Sorrells ME, Cisar G (1996) Prediction of heterosis in wheat using coefficient of parentage and RFLP-based estimates of genetic relationship. *Genome* 39: 1142-1149
- Beckmann JS, Soller M (1983) Restriction fragment length polymorphisms in genetic improvement: Methodologies, mapping and costs. *Theor Appl Genet* 67: 3543
- Bernardo R (1992) Relationship between single-cross performance and molecular marker heterozygosity. *Theor Appl Genet* 83: 628-634
- Bollich CN, Webb BD, Marchetti MA, Scott J (1991) Registration of 'Maybelle' rice. *Crop Sci* 31: 1090
- Boppenmaier J, Melchinger AE, Brunklaus-June E, Geiger HH, Herrmann RG (1992) Genetic diversity for RFLPs in European maize inbreds. I. Relation to performance of flint x dent crosses for forage traits. *Crop Sci* 32: 895-902
- Boppenmaier J, Melchinger AE, Seitz G, Gieger HH, Herrmann RG (1993) Genetic diversity of RFLP in European maize inbreds, III. Performance of crosses within versus between heterotic groups for grain traits. *Plant Breed* 111: 217-226
- Brar DS, Fujimura T, McCouch S, Zapata FJ (1994) Application of biotechnology in hybrid rice. In *Hybrid Rice Technology new developments and future prospects*. International Rice Research Conference, International Rice Research Institute, Manila, Philippines. pp 51-62
- Bruce AB (1910) The Mendelian theory of heredity and the augmentation of vigor. *Science* 32: 627-628
- Causse MA, Fulton TM, Cho YG, Ahn SN, Chunwongse J, Wu KS, Xiao JH, Yu ZH, Ronald PC, Harrington SE, Second G, McCouch SR, Tanksley SD (1994) Saturated molecular map of the rice genome based on an interspecific backcross population. *Genetics* 138: 1251-1274
- Cerna FJ, Cianzio SR, Rafalski A, Tingey S, Dyer D (1997) Relationship between seed yield heterosis and molecular marker heterozygosity in soybean. *Theor Appl Genet* 95: 460-467
- Charcosset A, Lefort-Buson M, Gallais A (1991) Relationship between heterosis and heterozygosity at marker loci: a theoretical computation. *Theor Appl Genet* 81: 571-575
- Cho YG, Eun MY, McCouch SR, Chae YA (1994) The semidwarf gene *sd-1*, of rice (*Oryza sativa* L.). II. Molecular mapping marker-assisted selection. *Theor Appl Genet* 89: 54-59
- Chowdari KV, Davierwala AP, Gupta VS, Ranjekar PK, Govila OP (1998a) Genotype identification and assessment of genetic relationships in pearl millet [*Penisetum glaucum* (L.) R. Br.] using microsatellites and RAPDs. *Theor Appl Genet* 97:154-162
- Chowdari KV, Venkatachalam SR, Davierwala AP, Gupta VS, Ranjekar PK, Govila OP (1998b) Hybrid performance and genetic distance as revealed by the (GATA)₄ microsatellite and RAPD markers in pearl millet. *Theor Appl Genet* 97:163-169
- Collins FS (1992) Positional cloning, let's not call it reverse anymore. *Nat Genet* 1: 3-6
- Crow JF (1952) Dominance and overdominance. In: Gowen JW (ed) *Heterosis*. Iowa State College Press, Ames. pp 282-297
- Darvasi A, Soller M (1994) Optimum spacing of genetic markers for determining linkage between molecular marker loci and quantitative trait loci. *Theor Appl Genet* 89: 351-357

- Davenport CB (1908) Degeneration, albinism and inbreeding. *Science* 28: 454-455
- Deng H, Wang G (1984) A study on prediction of heterosis in crops. *Hunan Agric Sci* 3: 1-5
- Devos KM, Millan T, Gale MD (1993a) Comparative RFLP maps of the homoeologous group-2 chromosomes of wheat, rye and barley. *Theor Appl Genet* 85: 784-792
- Devos KM, Atkinson MD, Chinoy CN, Harcourt RL, Koeber RMD, Liu CJ, Masojc P, Xie DX, Gale MD (1993b) Chromosomal rearrangements in the rye genome relative to that of wheat. *Theor Appl Genet* 85: 673-680
- Devos KM, Gale MD (1997) Comparative genetics in the grasses. *Plant Mol Biol* 35: 3-15
- Diers BW, McVetty PBE, Osborn TC (1996) Relationship between heterosis and genetic distance based on restriction fragment length polymorphism markers in oilseed rape (*Brassica napus* L.). *Crop Sci* 36: 79-83
- Dilday RH, Mgonja MA, Amonsli SA, Collins FC, Wells BR (1990) Plant height vs. mesocotyl and coleoptile elongation in rice: linkage or pleiotropism? *Crop Sci* 30: 815-818
- Dingkuhn M, De Datta SK, Pamplona R, Javellana C, Schnier HF (1992) Effect of late season N-fertilization on photosynthesis and yield of transplanted and direct-seeded tropical flooded rice. 2. A canopy stratification study. *Field Crop Res* 28: 235-249
- Dobzhansky T (1950) Genetics of natural populations. XIX. Origin of heterosis through natural selection in populations of *Drosophila pseudobscura*. *Genetics* 35: 288-302
- Dong NV, Subudhi PK, Luong PN, Quang VD, Quy TD, Zheng HG, Wang B, Nguyen HT (2000) Molecular mapping of a rice gene conditioning thermosensitive genic male sterility using AFLP, RFLP and SSR techniques. *Theor Appl Genet* 100: 727-734
- Dudley JW (1993) Molecular markers in plant improvement: manipulation of genes affecting quantitative traits. *Crop Sci* 33: 660-668
- Dudley JW, Saghai Maroof MA, Ruffner GK (1991) Molecular markers and grouping of parents in corn breeding programs. *Crop Sci* 31: 718-722
- East EM (1908) Inbreeding in corn. *Rep Conn Agric Exp Stn* 1907: 41-428
- Edwards MD, Page NJ (1994) Evaluation of marker-assisted selection through computer simulation. *Theor Appl Genet* 88: 376-382
- Edwards MD, Stuber CW, Wendel JF (1987) Molecular facilitated investigations of quantitative trait loci in maize. I. Number, genomic distribution and types of gene action. *Genetics* 116: 113-125
- Falconer DS (1960) *Introduction to Quantitative Genetics*. Ronald Press Co, New York, USA
- Falconer (1981) *An Introduction to Quantitative Genetics*. 2nd Edition, Longman, London, UK.
- Fasoulas AC (1988) *The Honeycomb Methodology of Plant Breeding*. AC Fasoulas, Thessaloniki, Greece, ISBN 960-220-000-6
- Fatokun CA, Menancio-Hautea D, Danesh D, Young ND (1992) Evidence for orthologous seed weight genes in cowpea and mungbean based on RFLP mapping. *Genetics* 132: 841-846
- Flavell RB (1995) Plant biotechnology R&D – The next ten years. *Tibtech* 13: 313-319
- Godshalk EB, Lee M, Lamkey KR (1990) Relationship of restriction fragment length polymorphisms to single-cross hybrid performance of maize. *Theor Appl Genet* 80: 273-280
- Gonella JA, Peterson PA (1978) Isozyme relatedness of inbred lines of maize and performance of their hybrids. *Maydica* 23: 55-61
- Gu M, You A, Pan X (1993) Analysis on allelic relationship of wide compatibility genes among several WCVs (*Oryza sativa* L.). *Sci Agric Sin* 26: 13-21

- Gu WK, Weeden NF, Yu J, Wallace DH (1995) Large-scale, cost-effective screening PCR products in marker-assisted selection applications. *Theor Appl Genet* 91: 465-470
- Hallauer AR and Miranda JB (1988) Quantitative genetics in maize breeding. Iowa State University Press, Ames.
- Hargrove TR, Cabanilla VL, Coffman WR (1988) Twenty years of rice breeding: the role of semidwarf varieties in rice breeding for Asian farmers and the effects of cytoplasmic diversity. *BioScience* 38: 675-681
- Harrington SE, Blight HFJ, Park WD, Jones CA, McCouch SR (1997) Linkage mapping of starch branching enzyme III in rice (*Oryza sativa* L.) and prediction of location of orthologous genes in other grasses. *Theor Appl Genet* 94: 564-568
- He HH, Zhang ZG, and Yuan S (1987) Response on development and fertility changes induced by light under different temperature condition in PGMS. *J Wuhan Univ* 8(7): 93
- Heidrich-Sobrinho E, Cordeiro AR (1975) Codominant isozymic alleles as markers of genetic diversity correlated with heterosis in maize (*Zea mays*). *Theor Appl Genet* 46: 197-199
- Helentjaris T (1988) Use of RFLP analysis to identify genes involved in complex traits of agronomic importance. In: *Current Communications in Molecular Biology - Genetic Improvement of Agriculturally Important Crops*. Cold Spring Harbor Laboratory. pp 2730
- Helentjaris T, Shattuck-Eidens D (1987) A strategy for pinpointing and cloning major genes involved in quantitative traits. *Maize Genet Coop News Lett* 61: 88
- Helentjaris T, Heun M (1994) Analysis of traits with complex inheritance in maize using molecular markers. In: Freeling M, Walbot V (ed), *The Maize Handbook*. Springer Verlag, New York, NY. pp 509-513
- Herd RW (1991) Research priorities for rice biotechnology. In: Khush GS, Toenniessen GH (eds) *Rice Biotechnology*. CAB Int, Wallingford, UK. pp 19-45
- Heydecker W (1960) Can we measure seedling vigor? *Proc Int Seed Test Assoc* 25: 498-512
- Huang N, Angeles ER, Domingo J, Magpantay G, Singh S, Zhang G, Kumaravadivel N, Bennett J, Khush GS (1997) Pyramiding bacterial blight resistance genes in rice: marker assisted selection using RFLP and PCR. *Theor Appl Genet* 95: 313-320
- Hunter RB, Kannenberg LW (1971) Isozyme characterization of corn (*Zea mays*) inbreds and its relationship to single cross hybrid performance. *Can J Genet Cytol* 13: 649-655
- Ikeda R, Wakasa K (1997) Frontier of rice breeding by utilization of genetic resources and biotechnology. In: Watanabe KN, Pehu E (ed) *Plant Biotechnology and Plant Genetic Resources for Sustainability and Productivity*. RG Landes Company, Austin. pp 103-115
- Ikehashi H (1991) Genetics of hybrid sterility in wide hybridization in rice (*Oryza sativa* L.). In: Y.S.P. Bajaj (ed) *Biotechnology in Agriculture and Forestry*. 14. Rice. Springer-Verlag, Berlin. pp 113-127
- Ikehashi H, Araki H (1984) Variety screening of compatibility types revealed in F1 fertility of distant cross in rice. *Jpn J Breed* 34: 304-313
- Ikehashi H, Araki H (1986) Genetics of F1 sterility in remote crosses of rice. In: International Rice Research Institute (ed) *Rice Genetics*. International Rice Research Institute, Manila, Philippines. pp 119-130
- Ikehashi H, Araki H (1988) Multiple alleles controlling F1 sterility in remote crosses of rice (*Oryza sativa*). *Jpn J Breed* 38: 283-291
- Ikehashi H, Zou JS, Hunn PM, Maruyama K (1994) Wide compatibility gene(s) and *indica japonica* heterosis in rice for temperate countries. In: Virmani SS (ed) *Hybrid Rice*

- Technology: New Developments and Future Prospects. International Rice Research Conference, International Rice Research Institute, Manila, Philippines. pp 21-31
- Iwabuchi M, Kyozuka J, Shimamoto K (1993) Processing followed by complete editing of an altered mitochondrial *atp σ* RNA restores fertility of cytoplasmic male sterile rice. *EMBO J* 12: 1437-1446
- Jena KK, Khush GS, Kochert G (1994) Comparative RFLP mapping of a wild rice, *Oryza officinalis* and cultivated rice, *O sativa*. *Genome* 37: 382-389
- Johnson GR (1991) RFLP assisted early generation selection for advanced generation testcross performance. Proc Twenty-seventh Illinois corn breeder's school, 4-5 March 1991, Campaigne, Illinois.
- Jones DF (1917) Dominance of linked factors as a means of accounting for heterosis. *Genetics* 2: 466-479
- Jones DE (1952) Plasmagenes and chromogenes in heterosis. In: Gowen JW (ed) *Heterosis*. Iowa State College Press, Ames, IA, USA. pp 224-235
- Jones DB, Peterson ML (1976) Rice seedling vigor at sub-optimal temperatures. *Crop Sci* 16: 102-105
- Joshi SP, Ranjekar PK, Gupta VS (1999) Molecular markers in plant genome analysis. *Curr Sci* 77: 230-240
- Joshi SP, Bhawe SG, Chowdari KV, Apte GS, Dhonukshe BL, Lalitha K, Ranjekar PK, Gupta VS (2001) Use of DNA markers in prediction of hybrid performance and heterosis for a three line hybrid system in rice. *Biochem Genet* (In Press)
- Kadowaki K, Suzuki T, Kazama S (1990) A chimeric gene containing the 5' portion of *atp σ* is associated with cytoplasmic male-sterility of rice. *Mol Gen Genet* 224: 10-16
- Kato S, Kosaka H, Hara (1928) On the affinity of rice varieties as shown by fertility of hybrid plants. *Bull Sci Fac Agric Kyushu Univ* 3: 132-147
- Katsuo K, Mizushima U (1958) Studies on the cytoplasmic difference among rice varieties, *Oryza sativa* L. *Jpn J Breed* 8: 1-5
- Keeble F, Pellew C (1910) The mode of inheritance of stature and flowering time in peas (*Pisum sativum*) *J Genet* 1: 47-56
- Khush GS, Kinoshita T (1991) Rice karyotype, marker genes, and linkage groups In: Khush GS, Toenniessen GH (eds) *Rice Biotechnology*. CAB Int, Wallingford, UK. pp 83-107
- Kinoshita T (1995) Report of the committee on gene symbolization, nomenclature and linkage groups. *Rice Genet Newsl* 12: 24-93
- Kinoshita T, Takahashi M (1991) The one hundredth report of genetical studies on rice plant—Linkage studies and future prospects. *J Fac Agric Hokkaido Univ* 65: 161
- Kumar LS (1999) DNA markers in plant improvement: An overview. *Biotechnology Advances* 17: 143-182
- Lamkey KR, Smith OS (1987) Performance and inbreeding depression of populations representing seven eras of maize breeding. *Crop Sci* 29: 1067-1071
- Lande R, Thompson R (1990) Efficiency of marker-assisted selection in the improvement of quantitative traits. *Genetics* 124: 743-756
- Lander ES, Botstein D (1989) Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics* 121: 185-199
- Lang NT, Subudhi PK, Virmani SS, Brar DS, Khush GS, Li Z, Huang N (1999) Development of PCR-based markers for thermosensitive genetic male sterility gene *tms3(t)* in rice. *Hereditas* 131: 121-127

- Larza LLB, de Souza CL Jr, Ottoboni LMM, Vieira MLC, de Souza AP. (1997) Genetic distance on inbred lines and prediction of maize single-cross performance using RAPD markers. *Theor Appl Genet* 94: 1023-1030
- Lee M, Godshalk EB, Lamkey KR, Woodman WW (1989) Association of restriction fragment length polymorphisms among maize inbreds with agronomic performance of their crosses. *Crop Sci* 29: 1067-1071
- Li CC, Rutger JN (1980) Inheritance of cool temperature seedling vigor in rice and its relationship with other agronomic characters. *Crop Sci* 20: 295-298
- Li Z, Zhu Y (1986) Rice male-sterile cytoplasm and fertility -restoration. In: Hybrid Rice. International Rice Research Institute, Manila, Philippines. pp 85-102
- Li Z, Zhu Y (1988) Rice male sterile cytoplasm and fertility restoration. In: Hybrid Rice. International Rice Research Institute, Manila, Philippines. pp 85-102
- Li H, Tang L, Zou J (1991) Marker-based analysis of wide compatibility in rice (in Chinese). *Hybrid Rice* 4: 22-24
- Li Z, Pinson SRM, Stansel JW, Park WD (1995) Identification of quantitative trait loci (QTLs) for heading date and plant height in cultivated rice (*Oryza sativa* L.) *Theor Appl Genet* 91: 374-381
- Li Z, Pinson SRM, Park WD, Paterson AH, Stansel JW (1997) Epistasis for three grain yield components in rice (*Oryza sativa* L.). *Genetics* 145: 453-465
- Li JZ, He P, Li SG, Lu RL, Zhu LH (2000a) Application of microsatellite markers for seed purity examination of a hybrid rice, Gangyou22. *Sheng Wu Kung Cheng Hsueh Pao* 16: 211-214
- Li JX, Yu SB, Xu CG, Tan YF, Gao YJ, Li LX, Zhang Q (2000b) Analyzing quantitative trait loci for yield using a vegetatively replicated F2 population from a cross between the parents of an elite rice hybrid. *Theor Appl Genet* 101: 248-254
- Lin SC, Yuan LP (1980) Hybrid rice breeding in China. In: International Rice Research Institute (ed) Innovative Approaches to Rice Breeding. IRRI, Manila, Philippines. pp 35-51
- Liu A, Zhang Q, Li H (1992) Location of a gene for wide compatibility in the RFLP linkage map. *Rice Genet Newsl* 9: 134-136
- Liu KD, Zhou ZQ, Xu CG, Zhang Q, Saghai Maroof MA (1996) An analysis of hybrid sterility in rice using a diallel cross of 21 parents involving indica, japonica and wide compatibility varieties. *Euphytica* 90: 275-280
- Liu KD, Wang J, Li HB, Xu CG, Liu AM, Li XH, Zhang Q (1997) A genome-wide analysis of wide compatibility in rice and the precise location of the S₅ locus in the molecular map. *Theor Appl Genet* 95: 809-814
- Livini C, Ajmone-Marsan P, Melchinger AE, Messmer MM, Motto M (1992) Genetic diversity of maize inbred lines within and among heterotic groups revealed by RFLPs. *Theor Appl Genet* 84: 17-25
- Martin GB, Brommonschenkel SH, Chunwongse J, Frary A, Ganai MW, Spivey R, Wu T, Earle ED, Tanksley SD (1993) Map-based cloning of a protein kinase gene conferring disease resistance in tomato. *Science* 262: 1543-1552
- Martin JM, Talbert LE, Lanning SP, Blake NK (1995) Hybrid performance in wheat as related to parental diversity. *Crop Sci* 35: 104-108
- Maruyama K, Araki H, Kato H (1990a) Enhancement of outcrossing habits of rice plant by mutation breeding. In: Proceedings of the Gamma Field Symposium, Institute of Radiation Breeding, NIAR, Ministry of Agriculture, Forestry and Fisheries, Japan. pp 1122

- Maruyama K, Araki H, Kato H (1990b) Thermosensitive genic male sterility induced by irradiation. In: Proceedings of the Second International Rice Genetics Symposium, 14-18 May, International Rice Research Institute, Manila, Philippines.
- Maruyama K, Araki H, and Kato H (1992) Daylength-sensitive growth stages and inheritance of photoperiod-sensitive genic male sterility (PGMS) in a rice line (X88). *Jpn J Breed* 42 (Suppl 1)
- McKenzie KS, Johnson CW, Tseng ST, Oster JJ, Brandon DM (1994) Breeding improved rice cultivars for temperate regions - a case study. *Aust J Exp Agric* 34: 897-905
- Melchinger AE (1993) Use of RFLP markers for analysis of genetic relationships among breeding materials and prediction of hybrid performance. In: Buxton DR, Shibles R, Forsberg RA, Blad BL, Asay KH, Paulson GM, Wilson RF (ed) *International Crop Science I*. Crop Science Society of America, Madison, WI, USA. pp 621-628
- Melchinger AE, Lee M, Lamkey KR, Woodman WL (1990a) Genetic diversity for restriction fragment length polymorphisms: Relation to estimated genetic effects in maize inbreds. *Crop Sci* 30: 1033-1040
- Melchinger AE, Lee M, Lamkey KR, Hallauer AR, Woodman WL (1990b) Genetic diversity of restriction fragment length polymorphisms and heterosis for two diallel sets of maize inbreds. *Theor Appl Genet* 80: 488-496
- Melchinger AE, Messmer MM, Lee M, Woodman WL, Lamkey KR (1991) Diversity and relationships among US maize inbreds revealed by restriction fragment length polymorphism. *Crop Sci* 31: 669-678
- Melchinger AE, Boppenmaier J, Dhillon BS, Pollmer WG, Herrmann RG (1992) Genetic diversity for RFLPs in European maize inbreds. II. Relation to performance of hybrids within versus between heterotic groups for forage traits. *Theor Appl Genet* 84: 672-681
- Messmer MM, Melchinger AE, Herrmann RG, Boppenmaier J (1993) Relationships among early European maize inbreds: II. Comparison of pedigree and RFLP data. *Crop Sci* 33: 944-950
- Mgonja MA, Ladeinde TAO, Akenova ME (1993) Genetic analysis of mesocotyl length and its relationship with other agronomic characters in rice (*Oryza sativa* L.). *Euphytica* 72: 189-195
- Minvielle F (1987) Dominance is not necessary for heterosis: a two-locus model. *Genet Res* 49: 245-247
- Moll RH, Lonnquist JH, Velez J, Johnson E (1965) The relationship of heterosis and genetic divergence in maize. *Genetics* 52: 139-144
- Morgante M, Rafalski A, Biddle P, Tingey S, Olivieri AM (1994) Genetic mapping and variability of seven soybean simple sequence repeat loci. *Genome* 37: 763-769
- Moser H, Lee M (1994) RFLP variation of genealogical distance, multivariate distance, heterosis and genetic variation in oats. *Theor Appl Genet* 87: 947-956
- Newton KJ (1988) Plant mitochondrial genomes: organization, expression and variation. *Ann Rev Plant Physiol Plant Mol Biol* 39: 503-532
- Otsuoka Y, Eberhart SA, Russell WA (1972) Comparisons of prediction formulas for maize hybrids. *Crop Sci* 12: 325-331
- Paterson AH, Tanksley SD, Sorrells ME (1991) DNA markers in plant improvement. *Advances in Agronomy* 46: 39-90
- Paterson AH (1995) Molecular dissection of quantitative traits: Progress and Prospects. *Genome Research* 5: 321-333

- Peng JY, Glaszman JC, Virmani SS (1988) Heterosis and isozyme divergence in indicarice. *Crop Sci* 28: 561-563
- Pereira MG, Lee M (1995) Identification of genomic regions affecting plant height in sorghum and maize. *Theor Appl Genet* 90: 380-388
- Pereira MG, Lee M, Barmel-Cox P, Woodman W, Doebley J, Whitkus R (1994) Construction of an RFLP map in sorghum and comparative mapping in maize. *Genome* 37: 236-243
- Peterson ML, Jones DB, Rutger JN (1978) Cool temperature screening of rice lines for seedling vigor. *II Riso* 27: 269-274
- Ranjhan S, Litts JC, Foolad MR, Rodriguez RL (1991) Chromosomal localization and genomic organization of α -amylase genes in rice (*Oryza sativa* L.). *Theor Appl Genet* 82: 481-488
- Reddy OUK, Siddiq EA, Sarma NP, Ali J, Hussain AJ, Nimmakayala P, Ramasamy P, Pammi S, Reddy AS (2000) Genetic analysis of temperature-sensitive male sterility in rice. *Theor Appl Genet* 100: 794-801
- Redona ED, Mackill DJ (1996a) Mapping quantitative trait loci for seedling vigor in rice using RFLPs. *Theor Appl Genet* 92: 395-402
- Redona ED, Mackill DJ (1996b) Molecular mapping of quantitative trait loci in *japonica* rice. *Genome* 39: 395-403
- Rice Research Group of Mian Yang Agricultural Institute of Sichuan Province (1978) An analysis on heterosis of F1 hybrids in rice. *Acta Sin Genet* 5: 158-164
- Russell WA, Eberhart SA (1970) Effects of three gene loci in the inheritance of quantitative characters in maize. *Crop Sci* 10: 165-169
- Russell WA, Eberhart SA, Urbano A, Vega O (1973) Recurrent selection for specific combining ability for yield in two maize populations. *Crop Sci* 13: 257-261
- Rutger JN (1984) Induced semidwarf mutants in rice. *Rice Genet Newsl* 1: 92
- Rutger JN, Bollich CN (1989) The use of introduced germoplasm in rice improvement. In: Shands HL, Weisner LE (eds) *Crop Science Society of America Special Publ No 7*. Crop Science Society America, Madison, USA. pp 1-13
- Saghai Maroof MA, Yang GP, Zhang Q, Gravois KA (1997) Correlation between molecular marker distance and hybrid performance in US southern long grain rice. *Crop Sci* 37: 145-150
- Sant VJ, Patankar AG, Gupta VS, Sarode ND, Mhase LB, Sainani MN, Deshmukh RB, Ranjekar PK (1999) Potential of DNA markers in detecting divergence and in analyzing heterosis in Indian elite chickpea cultivars. *Theor Appl Genet* 98: 1217-1225
- Schnell FW, Cockerham CC (1992) Multiplicative vs. arbitrary gene action in heterosis. *Genetics* 131: 461-469
- Shi MS (1981) Preliminary report of breeding and utilization of late japonica natural double-purpose line. *J Hubei Agric* 7: 1-3
- Shi MS (1985) The discovery and study of the photosensitive recessive malesterile rice (*Oryza sativa* L. subsp *japonica*). *Sci Agric Sin* 2: 44-48
- Shi MS, Deng JY (1986) The discovery, determination and utilization of the Hubei photosensitive genic male-sterile rice (*Oryza sativa* subsp *japonica*). *Acta Genet Sin* 13(2): 107-112
- Shinjo C (1975) Genetical studies of cytoplasmic male sterility and fertility restoration in rice, *Oryza sativa* L. *Sci Bull Coll Agric Univ Ryukyus* 22: 1-51
- Shinjo C (1984) Genetical studies of cytoplasmic male sterility and fertility restoration in rice having Genome A. *Biology of rice*. Japan Science Societies, Elsevier Press.

- Shoemaker RC, Lorenzen LL, Diers BW, Olson TC (1994) Genome mapping and agriculture. In: PM Greshoff (ed) Plant genome analysis. CRC Press, Boca Raton, Florida, USA. pp 1-10
- Shull GH (1908) The composition of a field of maize. Rep Am Breed Assoc 4: 296-301
- Shull GH (1911) The genotypes of maize. Am Nat 45: 234-252
- Shull GH (1914) Duplicate genes for capsule form in *Bursa bursa Bastoris*. J Ind Abst Vererb 12: 97-149
- Siddiq EA, Jachuck PJ, Mahadevappa M, Zaman FU, Vijaya Kumar, Vidyachandra B, Sidhu GS, Ish Kumar, Prasad MN, Rangaswamy M, Pandey MP, Panwar DVS, Ilyas Ahmed (1994) Hybrid rice research in India. In: Virmani SS (ed) Hybrid Rice Technology: New Developments and Future Prospects. International Rice Research Conference, International Rice Research Institute, Manila, Philippines. pp 157-171
- Singh SD, Singh P, Rai KN, Andrews DJ (1990) Registration of ICMA841 and ICMH841 pearl millet parental lines with aa cytoplasmic-genic male sterility system. Crop Sci 30: 1378
- Smith JS, Smith OS (1989) The description and assessment of distances between inbred lines of maize. II. The utility of morphological, biochemical and genetic descriptors and a scheme for the testing of distinctness between inbred lines. Maydica 34: 151-161
- Smith JS, Smith OS (1991) Restriction fragment length polymorphisms can differentiate among US maize hybrids. Crop Sci 31: 893-899
- Song WY, Wang GP, Chen LL, Kim HS, Pi LY, Holsten T, Gardner J, Wang B, Zhai WX, Zhu LH, Fauquet C, Ronald P (1995) A receptor kinase-like protein encoded by the rice disease resistance gene, *Xa21*. Science 270: 1804-1806.
- Sonti RV, Yashitola Y, Thirumurgan T, Sundaram RM, Ramesha MS, Sarma NP (2000) DNA markers for assessment of genetic purity of rice hybrids. In: 4th Institutional Rice Genetics Symposium, 22-27 October, 2000, International Rice Research Institute, Manila, Philippines.
- Sprague GE (1983) Heterosis in maize: Theory and practice. In: R Frankel (ed) Heterosis: Reappraisal of Theory and Practice. Springer-Verlag, New York, NY, USA. pp 47-70
- Sprague GE, Miller PA (1950) A suggestion for evaluating current concepts of the genetic mechanism of heterosis in corn. Agron J 42: 161-162
- Staub JE, Kuhns LJ, May B, Grun P. (1982) Stability of potato tuber isozymes under different storage regimes. J Am Sci 107: 405-408
- Stuber CW (1995) Mapping and manipulating quantitative traits in maize. Trends Genet 11: 477-481
- Stuber CW, Edwards MD (1986) Genotypic selection for improvement of quantitative traits in corn using molecular marker loci. In: Proceedings 41st Annual Corn and Sorghum Research Conference, American Trade Seed Association, Washington, DC, USA. pp 70-83
- Stuber CW, Williams WP, Moll RH (1973) Epistasis in maize (*Zea mays* L.). III. Significance in predictions of hybrid performances. Crop Sci 13: 195-200
- Stuber CW, Lincoln SE, Welff DW, Helentjaris T, Lander EC (1992) Identification of genetic factors contributing to heterosis in a hybrid from two elite maize inbred lines using molecular markers. Genetics 132: 823-839
- Sun ZX, Xiong ZM, Min SK, Si HM (1989) Identification of the temperature-sensitive male sterile rice. Chin J Rice Sci 3 (2): 49-55
- Tanksley SD (1993) Mapping polygenes. Annu Rev Genet 27: 205-233

- Tanksley SD, Nelson JC (1996) Advanced backcross QTL analysis: a method for the simultaneous discovery and transfer of valuable QTLs from unadapted germplasm into elite breeding lines. *Theor Appl Genet* 92: 191-203
- Tanksley SD, Bernatzky R, Lapitan NL, Price JP (1988) Conservation of gene repertoire but not gene order in pepper and tomato. *Proc Natl Acad Sci, USA* 85: 6419-6423
- Tanksley SD, Young ND, Paterson AH, Bonierbale MW (1989) RFLP mapping in plant breeding: New tools for an old science. *Bio/Technology* 7: 257-264
- Tanksley S, Causse M, Fuslton T, Ahn N, Wang Z, Wu K, Xiao J, Yu Z, Second G, McCouch S. (1992) A high density molecular map of rice genome. *Rice Genet Newsl* 9: 111-115
- Tanksley SD, Grandillo S, Fulton TM, Zamir D, Eshed Y, Petiard V, Lopez J, Beck-Bunn T (1996) Advanced backcross QTL analysis in a cross between an elite processing line of tomato and its wild relative *L. pimpinellifolium*. *Theor Appl Genet* 92: 213-224
- Teutonico RA, Osborn TC (1994) Mapping of RFLP and quantitative trait loci in *Brassica rapa* and comparison to the linkage maps of *B. napus*, *B. oleracea* and *Arabidopsis thaliana*. *Theor Appl Genet* 89: 885-893
- Thomas BR, Rodriguez RL (1994) Metabolite signals regulate gene expression and source/sink relations in cereal seedlings. *Plant Physiol* 106: 1235-1239
- Tsiftaris SA (1995) Molecular aspects of heterosis in plants. *Physiologia Plantarum* 94: 362-370.
- Tsiftaris AS, Efthimiadis P (1987) F1 heterosis and heterozygosity for isozymic structural loci in maize. In: Rattazi MC, Scandalios JC, Whitt GS (eds) *Isozymes: Current Topics in Biological and Medical Research*, vol 16. Alan R Liss, New York, NY, USA. pp 157-174
- Turner FT, Chen CC, Bollich CN (1982) Coleoptile and mesocotyl length in semidwarf rice seedlings. *Crop Sci* 22: 43-46
- Virmani SS (1992) Transfer and induction of thermosensitive genic male sterile mutant in indica rice. In: *Proceeding of the Second International Symposium on Hybrid Rice*, 21-25 Apr 1992. International Rice Research Institute, Manila, Philippines.
- Virmani SS, Aquino RC, Khush GS (1982) Heterosis breeding in rice. *Theor Appl Genet* 63: 373-380
- Virmani SS, Shinjyo C. (1988) Current status of analysis and symbols for male-sterile cytoplasm and fertility-restoring genes. *Rice Genet Newsl* 5: 9-15
- Virmani SS, Young JB, Moon HP, Kumar I, Flinn JC (1991) Increasing rice yields through exploitation of heterosis. *IRRI Res Pap Ser* 156. pp 13
- Vodenicharova M (1989) Use of proteins as molecular genetic markers in plants. *Genet Sel* 22: 269-277
- Walejko RN, Russell WA (1977) Evaluation of recurrent selection for specific combining ability in two open pollinated maize cultivars. *Crop Sci* 17: 647-651
- Wang GL, Wing RA, Paterson AH (1993) PCR amplification from single seeds, facilitating DNA marker-assisted breeding. *Nucl Acids Res* 21: 2527
- Wang B, Zheng J, Xu W, Wu W, Zheng H, Nguyen HT (1995a) Tagging and mapping of rice thermosensitive genic male sterile (TGMS) gene. *Plant Genome* III, January.
- Wang B, Xu WW, Wang JZ, Wu W, Zheng HG, Yang ZY, Ray JD, Nguyen HT (1995b) Tagging and mapping the thermo-sensitive genic male-sterile gene in rice (*Oryza sativa* L.) with molecular markers. *Theor Appl Genet* 91: 1111-1114
- Wang B, Fu J, Zheng H, Qiu F, Li C, Xie W, Wang J, Jin D, Yang R (1998) Construction of rice BAC libraries and screening with AFLP markers linked to the TGMS gene. *Plant and Animal Genome VI Conference*, San Diego, California, USA, January 18-22.

- Weeden NF, Timmerman M, Hermat M, Kneen BE, Lodhi MA (1992) Inheritance and reliability of RAPD markers. In: J Nienhuis (ed) Proc Symp applications of RAPD technology to Plant Breeding, Nov 1992, Minneapolis, Minnesota, USA. pp 1217
- Wicking E, Williamson B (1991) From linked marker to gene. Trends Genet 7: 288-293
- Williams JF, Peterson ML. (1973) relations between alpha-amylase activity and growth of rice seedlings. Crop Sci 13: 612-615
- Wu XJ, Yin HQ, Yin H (1991) Preliminary study of the temperature effect of Annong S1 and W6154 S. Crop Res (China) 5: 4-6
- Xiao J, Li J, Yuan L, Tanksley SD (1995) Dominance is the major basis of heterosis in rice as revealed by QTL analysis using molecular markers. Genetics 140: 745-754
- Xiao J, Li J, Yuan L, McCouch SR, Tanksley SD (1996) Genetic diversity and its relationship to hybrid performance and heterosis in rice as revealed by PCR based markers in rice Theor Appl Genet 92: 637-643
- Xiao J, Li J, Grandillo S, Ahn SN, Yuan L, Tanksley SD, McCouch SR (1998) Identification of trait improving quantitative trait loci alleles from a wild rice relative, *Oryza rufipogon*. Genetics: 150: 899-909
- Yanagihara S, McCouch SR, Ishikawa K, Ogi Y, Maruyama K, Ikehashi H (1995) Molecular analysis of the inheritance of the S-5 locus conferring wide compatibility in indica / japonica hybrids of rice (*O. sativa* L). Theor Appl Genet 90: 1824-88
- Yang YC, Li WM, Wang NY, Liang KJ, Chen QH (1989) Discovery and preliminary study of indica photosensitive genic male sterile germplasm 5460ps. Chin J Rice Sci 3(1): 47-48
- Yoshimura S, Yoshimura A, Iwata N, McCouch SR, Abenes ML, Baraoidan MR, Mew TW, Nelson RJ (1995a) Tagging and combining bacterial blight resistance genes in rice using RAPD and RFLP markers. Mol Breed 1: 375-387
- Yoshimura S, Umehara Y, Kurata N, Nagamura Y, Sasaki T, Minobe Y, Iwata N. (1996) Identification of a YAC clone carrying the *Xa-1* allele, a bacterial blight resistance gene in rice. Theor Appl Genet. 93: 117-122
- Young ND (1990) Potential applications of map-based cloning to plant pathology. Physiol Mol Plant Pathol 37: 81-94
- Young JB, Virmani SS (1990) heterosis in rice over environments. Euphytica 51: 87-93
- Yu, S.B., Li, J.X., Xu, C.G., Tan, Y.F., Gao, Y.J., Li, X.H., Zhang, Q., Saghai Maroof, M.A. (1997) Importance of epistasis as the genetic basis of heterosis in an elite rice hybrid. *Proc. Natl. Acad. Sci., USA*. 94, 9226-9231
- Yuan LP (1985) Exploration of super high yielding hybrid rice. Hybrid Rice 3: 1-3
- Yuan LP (1987) Strategy conception of hybrid rice breeding. Hybrid Rice 1: 13
- Yuan LP (1992a) Development and prospects of hybrid rice breeding. In: You C, Chen ZL (eds) Agricultural Biotechnology. Proceedings Asia-Pacific Conference on Agricultural Biotechnology, Beijing, August 20-24, 1992. pp 97-105
- Yuan LP (1992b) The strategy of breeding rice PGMS and TGMS lines. Hybrid Rice 1: 14
- Yuan LP, Cheng HX (1986) Hybrid rice breeding and cultivation. Hunan Science and Technology Press, Hunan, China.
- Zhai W, Lu C, Zhu L, Yang W, Zhang Q (1996) PCR analysis of halfseeds of cereal crops and its application in marker-assisted selection and breeding. Chin J Biotechnol 12: 249-255
- Zhang ZG, Yuan SC, (1989) Effects of twilight on two photoperiod reactions in Hubei photoperiod-sensitive genic male sterile rice PGMS Chi J Rice Sci 3(3): 107-112

- Zhang ZG, Yuan SC (1992) Studies on the conditions and adaptability of PGMS fertility alteration. In: Liu HL (ed) *Advances in Crop Cultivation and Breeding*. China Agricultural Publishing House, Beijing, China. pp 1-32
- Zhang ZG, Yuan SC, Xu CZ. (1987) The influence of photoperiod on the fertility changes of Hubei photoperiod - sensitive genic male sterile rice (PGMS) *Chin J Rice Sci* 1(3): 136-143
- Zhang ZG, Yuan SC, Zen HL, Li YZ, Zhang DP (1992a) Studies on the genetics of two photoreaction on photo-thermosensitive sterility. *J Huazhong Agric Univ* 11: 742
- Zhang ZG, Zeng HL, Yuan SC, Wang BX, Li YZ, Zhang DP (1992b) Studies on the model of photothermo reaction of fertility alteration in photosensitive genic male sterile rice. *J Huazhong Agric Univ* 11: 1-6
- Zhang QF, Gao YJ, Yang SH, Ragab RA, Saghai Maroof MA, Li ZB (1994a) A diallel analysis of heterosis in elite hybrid rice based on RFLPs and microsatellites. *Theor Appl Genet* 89: 185-192
- Zhang Q, Shen BZ, Dai XK, Mei MH, Saghai Maroof MA, Li ZB (1994b) Using bulked extremes and recessive class to map genes for photoperiodsensitive genic male sterility in rice. *Proc Natl Acad Sci USA* 91: 8675-8679
- Zhang QF, Gao YJ, Saghai Maroof MA, Yang SH, Li JX (1995) Molecular divergence and hybrid performance in rice. *Mol Breed* 1: 133-142
- Zhang Q, Zhou ZQ, Yang GP, Xu CG, Lin KD, Saghai Maroof MA (1996) Molecular marker heterozygosity and hybrid performance in indica and japonica rice. *Theor Appl Genet* 93: 1218-1224
- Zhang Q, Liu KD, Yang GP, Saghai Maroof MA, Xu CG, Zhou ZQ (1997) Molecular marker diversity and hybrid diversity in indica-japonica rice crosses. *Theor Appl Genet* 95: 112-118
- Zheng K, Shen P, Qian H, Wang J, (1992) Tagging genes for wide compatibility in rice via linkage to RFLP markers. *Chinese J Rice Sci* 6: 145-150
- Zhou TB, Xiao HC, Lei DY, Duan QX. (1988) The breeding of indica photosensitive male sterile line. *J Hunan Agric Sci* 6: 16-18
- Zhou GF, Gong GM, Yin CQ, Shen XB (1991) Preliminary observation on fertility alteration duration and genetics of Annong S-1, an indica two-usage line. *J Hunan Agric Sci (China)* 4: 10-11
- Zhuang CX, Zhang GQ, Mei MT, Lu YG (1999) Molecular mapping of the S-a locus for F1 pollen sterility in cultivated rice. *I Chuan Hsueh Pao* 26: 213-218

MOLECULAR MARKERS AND ABIOTIC STRESSES

Following the Signal Maze to Protective Metabolism in Abiotic Stress - Search for Opportune Shortcuts to Improve Resistance

I. WINICOV

Arizona State University

Department of Plant Biology, PO Box 871601, Tempe, AZ85287, USA

INTRODUCTION

In most moderate climates plants encounter a variety of abiotic stress conditions during the growing season which significantly impact their productivity. For purposes of this review *abiotic stress* will be defined as climatic and soil variation in osmotic environment (salt, drought and cold), daily or seasonal temperature variation (cold and heat) and oxidative stress exacerbating the other stress conditions especially in presence of high light intensity. These stress conditions can be of short duration, such as temperature variations during the day and night or the drying of fields. The stress can also be prolonged as during seasonal changes of heat and cold, while drought and salinity conditions are brought on by lack of sufficient water and poor irrigation practices. The goal in crop improvement is to extend plant tolerance to these stress conditions for survival and continued productivity under the adverse growing conditions.

Plants sense gradual changes in environment and can adapt in a limited way to those changes by activating genes in their genetic repertoire that allow them to acclimate to the altered environment, whether it be salinity/drought (Hasegawa et al., 2000; Shinozaki et al., 1999; Winicov, 1998), cold (Kingston-Smith et al., 1999; Thomashow, 1998), heat (Vierling, 1991) or high light and oxidative stress (Allen, 1995; Foyer et al., 1994). The acclimation process and the varietal differences in plant species resistance to different abiotic stress conditions indicate that judicious manipulation of expression of the genetic repertoire of plants could lead to improved stress resistance. The molecular approach has received considerable amount of attention, since traditional plant breeding has had very limited success in improving salt/drought, cold or heat tolerance of crop plants in the past.

Although many of the molecular changes that occur in plants during abiotic stress are known, the understanding on how to utilize this knowledge to engineer plants with improved tolerance to these stresses is still in the developmental stages. In part this is due to the magnitude of the changes in gene expression of plants exposed to stress since more and more gene transcripts are being identified as stress induced with the increased technology for measuring quantitative trait loci (QTL), expressed sequence tags (EST) and micro-array analyses (Kawasaki et al., 2002; Seki et al., 2001). While characterization to date has focused predominantly on genes induced rapidly by acute stress, little information is available on the expression of genes that are still induced after prolonged stress (Winicov, 2000b). In addition, salt, drought and cold tolerance appear to be quantitative traits in all species assessed this far.

The following discussion will provide examples of molecular strategies used to date to alleviate abiotic stress in plants, primarily based on overexpression of genes activated by stress in plants or found to be associated with tolerance in yeast and bacteria, with the assumption that they may constitute useful markers for tolerance in crop improvement. Most of these represent single gene manipulation in the transgenic plant. Recent focus has also shifted on the potential for improving tolerance by boosting multiple gene expression of integrated stress defense pathways by manipulation of signaling pathways or transcription factors that participate in stress gene regulation.

1. OSMOTIC STRESS MANAGEMENT

This section will discuss current attempts to increase tolerance in plants to salinity, drought and cold. Salinity and drought contribute to much of the yield reduction in agriculture throughout the world. Overcoming salinity and drought stress, both of which lead to cellular osmotic and oxidative problems therefore is a challenge to provide sustainable agriculture for the burgeoning world population. The associated problems of soil salinity, physiology, osmotic effects and nutritional imbalance have been recently reviewed (Jain and Selvaraj, 1997). Cold temperatures near freezing also restrict agricultural potential by severely disrupting membrane systems (Steponkus, 1984) of the cell that lead to dehydration causing plant damage and death. This suggests that mechanisms of acclimation to freezing temperatures are likely to induce at least some cellular components similar to

those that protect against changes in osmotic potential caused by salinity and drought and is supported by current experimental findings. There is also increasing evidence that the signal transduction pathways for inducing expression of salt, drought and cold-regulated genes may utilize similar protein kinase cascades and calcium as an important second messenger (Shinozaki and Yamaguchi-Shinozaki, 1997; Thomashow, 1999).

Crop yield, especially under osmotic stress, is a polygenic trait, and mapping of salt/drought and cold stress affected genes has uncovered a highly complex picture. It was hoped that the advent of new techniques in germplasm characterization with RFLP tags for genomic regions associated with desirable traits and marker assisted selection of QTLs would speed up breeding crop plants with improved salt/drought tolerance. However, the results to date have been less useful than initially predicted (Ribaut and Hoisington, 1998). An other strategy for salt stress resistance improvement in crop plants was initially developed in our laboratory for alfalfa and long grained California rice by selection in tissue culture (Winicov, 1991; Winicov, 1996), but the mutation(s) that provided the heritable improvement in salt tolerance have not been identified at the molecular level. Transgenic plants have been engineered in a number of laboratories to express individual genes of plant, yeast or bacterial origin normally induced under salt/drought stress, in order to test their efficacy for improving salt/drought and cold tolerance. The results from these tests for resistance due to single gene transfer have been mixed, indicating the need to address the question of multigenic responses necessary for significant improvement in tolerance that could be sustained over long- term growth in the field.

1.1. PHENOTYPE TARGETED BREEDING/SELECTION FOR RESISTANCE

Different chromosomal regions have been identified by QTL mapping in a variety of plants to be associated with ability to provide increased osmotic adjustment in drought or salt stress (Foolad and Jones, 1993; Galiba et al., 1992; Lebreton et al., 1995; Lilley et al., 1996; Teulat et al., 1998). While some of the mapped regions overlap between different plants, different genes may be involved in the tolerance phenotype in different plants. Many of these dehydration stress genes appear to belong to large gene families with a broad chromosomal distribution as demonstrated by the wheat dehydrins (Werner-Fraczek and Close, 1998). Interestingly, one of these dehydrin

alleles has been linked with chilling resistance during seedling emergence (Ismail et al., 1999).

Although heritability of water-use efficiency trait has been reported in alfalfa using selection based on molecular marker diversity (Ray, 1999), other results with inbred lines of rice with amplified fragment length polymorphism (AFPL) failed to show marker association with salinity tolerance (Flowers et al., 2000). While mapping experiments with glycinebetaine accumulation in maize (Yang et al., 1995) appear promising, evidence for a general pattern in marker distribution is still lacking.

Cold acclimation in higher plants also has been found to be a polygenic trait (Thomashow, 1990) and most chromosomes can contribute a measure of tolerance. In wheat, the interval *Vrn1-Fr1* on chromosome 5A, associated with vernalization, has been identified as the region important in freezing tolerance (Galiba et al., 1995) and appears to contain a number of genes important for tolerance (Storlie et al., 1998), although the mechanism provided for tolerance from this interval remains to be determined. New marker assisted breeding strategies will continue to emerge, but are likely to require development of near-isogenic lines and screening of large number of plants to provide unambiguous data.

The alternate strategy for salt stress resistance improvement in crop plants from our laboratory utilizes the ability to select cell lines in callus culture at mutational frequencies, that have acquired the ability to grow at previously lethal concentrations (172 mM NaCl) of salt (Winicov et al., 1989). These cell lines regenerated fertile plants with heritably improved salt tolerance compared to the parent plants from which the cells were derived or plants regenerated from culture without the NaCl selection (Winicov, 1991; Winicov, 1996). The salt-tolerance trait was semi-dominant in both alfalfa and rice regenerated from tissue culture. The stability of the selected mutation was shown recently in experiments that re-tested the improved salt tolerance of the original regenerated plants, which had been maintained by cuttings for nearly ten years (Winicov and Bastola, 1999). These results demonstrated that cellular tolerance in both alfalfa and rice could be utilized at the whole plant level and is a viable method for obtaining plants with improved salt tolerance. Although we have identified and cloned a number of genes that are up-regulated and salt inducible in the salt tolerant alfalfa (Deutch and Winicov, 1995; Winicov, 1993; Winicov and Button, 1991;

Winicov and Krishnan, 1996; Winicov and Shirzadegan, 1997), we have not identified the specific semi-dominant mutation responsible for enhanced salt-tolerance that is associated with this multigenic change in gene expression.

1.2. IDENTIFICATION OF POTENTIAL MARKERS BY MUTATIONAL ANALYSIS

Mutational approach in *Arabidopsis* has targeted identification of genes that would provide potential markers for salt/drought or cold tolerance by selection of seedlings for mutations that display a “hypersensitive” response to the applied stress or “constitutive” tolerance relative to the wild type plants. The hypersensitivity assay to NaCl has identified mutants that are more sensitive to salt injury than wild type plants. These represent three loci in *Arabidopsis* that are involved in essential signaling processes necessary to maintain K⁺ nutrition at levels adequate for average NaCl concentrations and indicates that interruption of calcium signaling through a calcineurin-like pathway mediates the hypersensitivity response (Liu and Zhu, 1997a; Zhu et al., 1998b). Subsequent cloning of the *SOS2* and the *SOS3* genes selected by this screen and their characterization suggests that resistance to hypersensitivity may be mediated by the kinase/calcium binding protein complex (Halfter et al., 2000; Liu et al., 2000; Liu and Zhu, 1998) and involve a Na⁺/H⁺ antiporter (Shi et al., 2000). It will be interesting to see if manipulation of this pathway can improve salt tolerance above the average limits observed for individual plant species.

Mutants impaired in freezing tolerance were identified by selection of seedlings from chemically mutagenized plants that did not show a chilling-sensitive phenotype during the cold acclimation treatment, but nevertheless did not attain normal levels of freezing tolerance (McKown et al., 1996; Warren et al., 1996). The five *sfr* mutations showed developmental differences in expression and while some appeared to decrease the cryostability of the plasma membrane, others showed a variety of metabolic changes. Constitutive freezing tolerance has been shown in mutants of the *eskimo1* (*esk1*) gene, that require no prior acclimation for normal levels of cold-tolerance and showed slightly increased freezing-tolerance after acclimation (Xin and Browse, 1998). The *esk1* two recessive alleles increased the expression of some genes associated with cold acclimation under normal growth conditions, but not others, suggesting that ESK1 may act as a negative regulator. Another negative regulator of low

temperature signal transduction appears to be encoded by the *HOS1* gene in *Arabidopsis* (Ishitani et al., 1998). However, since the *hos1* mutation also results in decreased stimulation of gene expression by ABA, high salt and polyethylene glycol of some of the same genes super-induced by cold (Ishitani et al., 1997), the HOS1 protein may also have a positive role in gene induction by ABA and osmotic stress, thus providing a point of differentiation between cold and osmotic stress. The mechanism of action for most of these mutations is not known, but the pleiotropic effects exhibited by these mutations emphasize the inter-connected balance of each of the signaling and metabolic pathways involved in these stress responses.

1.3. TRANSGENIC APPROACH TO INCREASE RESISTANCE TO OSMOTIC STRESS-SINGLE GENE MANIPULATION

It is now well established that salt, drought and cold stress result in extensive changes in gene expression with induction of many genes. Current information related to osmotic stress gene induction and the molecular mechanisms of tolerance in bacteria and yeast have been recently reviewed (Bohnert and Sheveleva, 1998; Ingram and Bartels, 1996; Shinozaki and Yamaguchi-Shinozaki, 1997; Thomashow, 1999). An increasing number of genes are being tested in transgenic plants for their ability to provide improved resistance to salt, drought and cold stress (Thomashow, 1999; Winicov, 2000b). The predominant effort has been to engineer increased levels of compatible solutes or osmoprotectants as well as proteins that may stabilize macromolecules under conditions of decreasing water activity. Since oxidative stress injury has been considered to be a common component of abiotic stress damage in plants (Noctor and Foyer, 1998) a significant number of antioxidant enzymes also have been cloned and expressed in transgenic plants to test their role as protectants to abiotic stress especially in relationship to high light and its resulting damage to the photosystems.

1.3.1. Overexpression of Osmoprotectants

Compatible solutes or osmoprotectants occur in all organisms, including plants and serve to increase osmotic pressure in the cell as well as stabilize proteins and membranes under abiotic stress conditions. Osmoprotectants can be sugars, polyols, betaines or amino acids such as proline, that have been shown to accumulate in plant adaptation to salt, drought and cold

stress. Recently the discovery of 3-dimethylsulfoniopropionate (DMSP) in marine algae (Gage et al., 1997) and plants (Trossat et al., 1998) has potential for osmoprotectant production not requiring nitrogen. In addition, it has been possible to engineer ectoine expression in tobacco cells with transgenic expression of three genes from *Halomonas elongata* (Nakayama et al., 2000) and improve salt tolerance of the cells. Therefore, increasing the cellular concentration of these compounds has been a target for metabolic engineering in plants, with genes encoding enzymes for their production from plants, bacteria or yeast. Table 1 summarizes the data from different laboratories where genes for osmoprotectant production have been introduced in transgenic plants.

The incremental improvement in stress tolerance of the transgenic plants however was accompanied in some cases by pleiotropic effects on growth resulting in dwarfed phenotypes and in the case of sorbitol production necrotic lesions on the leaves of plants producing highest levels of sorbitol. Our understanding of the metabolic balance required for transgenic plants to overproduce sufficient levels of the desired osmoprotectant is still incomplete nor do we know the actual molecular mechanisms affording protection from stress. It has been suggested that mannitol may exert its protective effects through increasing resistance to oxidative stress through radical scavenging mechanisms, since the cellular mannitol concentrations in transgenic plants are too low to provide for significant osmotic adjustment (Shen et al., 1997). Other metabolic insufficiencies arise where the introduced transgenic enzyme is substrate limited by endogenous metabolic pathways as in the case of *myo*-inositol for D-ononitol production (Sheveleva et al., 1997) or choline supply for Gly-betaine production (Nuccio et al., 1999; Nuccio et al., 1998). Additional factors could become limiting in continued recycling of S-adenosylhomocysteine (SAH), a potent inhibitor of S-adenosylmethionine dependent methylation reactions that are required for choline precursor synthesis. SAH catabolism depends on SAH hydrolase and adenosine kinase, both salt inducible enzymes in glycine betaine accumulating plants, but not in non-accumulating plant species (Dr.

TABLE 1. Metabolically engineered plants to overexpress compatible solutes for protection against osmotic stress.

SOLUTE	PLANT	GENE, Source	STRESS	PLEOTROPIC SIDE - EFFECTS	REFERENCE
Fructan	tobacco	<i>SacB</i> , bacteria	S	No	(Pilon-Smits et al., 1995)
Trehalose	tobacco	<i>Tps1</i> , yeast	S	Yes	(Holmstrom et al., 1996)
Trehalose	tobacco	<i>Tps1</i> , yeast	D	Yes	Romero, 1997 #806]
Mannitol	tobacco	<i>mtlD</i> , bacteria	S	No	(Tarczynski et al., 1993)
Mannitol	tobacco	<i>mtlD</i> , bacteria	S	Yes	(Karakas et al., 1997)
Mannitol	Arabidopsis	<i>mtlD</i> , bacteria	S	No	(Thomas et al., 1995)
D-ononitol	tobacco	<i>Imt1</i> , plant	S, D	No	(Sheveleva et al., 1997)
Sorbitol	tobacco	<i>Spd1</i> , plant	S	Yes	(Sheveleva et al., 1998)
Proline	tobacco	<i>P5CS</i> , plant	S	No	(Kishor et al., 1995)
Proline	rice	<i>P5CS</i> , plant	S, D	No	(Zhu et al., 1998a)
Proline	Arabidopsis	antisense- <i>proDH</i>	S, C		(Nanjo et al., 1999a)
Gly-betaine	Arabidopsis	<i>codA</i> , bacteria	S, C, H, L	No	(Alia et al., 1999)
Gly-betaine	tobacco	<i>cmo</i> , plant	S		(Nuccio et al., 1998)
Gly-betaine	tobacco	<i>betL4</i> , bacteria	S		(Lilius et al., 1996)
Gly-betaine	rice	<i>codA</i> , bacteria	S, C		(Sakamoto et al., 1998)

Stress: Detectable improvement in response to: S, salt; D, drought; L, light; H, heat; C, cold or freezing stress.

E. Weretilynk, personal communication).

Increased levels of proline as an osmoprotectant have been engineered by increased proline synthesis (Kishor et al., 1995) and decreased proline degradation by antisense methods (Nanjo et al., 1999a). The true role for proline in stress protection however, remains to be clarified (Verma, 1999), since acquired salt-tolerance in alfalfa was not associated with significant changes in proline accumulation, except in a brief temporal manner in the roots (Petrusa and Winicov, 1997) and increased proline concentrations did not protect salt hypersensitive *Arabidopsis* against salt stress (Liu and Zhu, 1997b). While it has been shown that proline is essential for plants under conditions of osmotic stress and for morphological development (Nanjo et al., 1999b), very high levels of proline could lead to proline hypersensitivity through production of toxic proline catabolites and impact carbohydrate metabolism (Hellmann et al., 2000).

1.3.2. *Overexpression of gene products that affect membrane function and/or stability.*

Ion uptake, transfer across membranes and compartmentalization in the plant are an integral metabolic change in salt/drought stress and adaptation (Niu et al., 1995). Freezing leads to membrane damage and subsequent dehydration and electrolyte leakage (Steponkus et al., 1993). Phospholipid metabolism is intricately involved in membrane reorganization, metabolic channeling of fatty acids and is a known source of second messengers (Chapman, 1998) and thus would be expected to play a significant role in adaptation to abiotic stress. In fact, phospholipase D has been implicated in the early events leading to dessication survival in *Craterostigma plantagineum* (Frank et al., 2000). Increased levels of trienoic fatty acids have been associated with cold acclimation of higher plants and overexpression of chloroplast ω -3 fatty acid desaturase in transgenic tobacco gave some protection against chilling-induced damage (Kodama et al., 1994). Also, freeze-induced reactive oxygen species cause significant membrane damage (McKersie and Bowley, 1997), indicating that enhancement of membrane function and stability could play a significant role in improving osmotic tolerance.

Genes encoding membrane components of ion pumps could be targets for activation to overcome salt stress. This was demonstrated in yeast, where a

mutation in the trans-membrane domain of the high affinity K^+ -transporter, HKT1, allowed for increased Na^+ tolerance (Rubio et al., 1996). The similarity of transport determinants between yeast and plants (Serrano et al., 1999) is supported by identification of *Arabidopsis* mutants hypersensitive to NaCl that are deficient in the high affinity K^+ transport system (Liu and Zhu, 1997a). The recent introduction of a cDNA encoding an *Arabidopsis* Na^+/K^+ antiport, normally localized in the prevacuolar compartment, has shown that overexpression of this gene in *Arabidopsis* can increase salt tolerance of the transgenic plants as measured under a regimen of adaptation to salt (Apse et al., 1999). Even though the endogenous gene induction by salt was undetectable, the modest overexpression of this antiport provided survival to salt stress. Overexpression of the same gene in tomato significantly improved salinity tolerance of the transgenic plants grown hydroponically (Zhang and Blumwald, 2001).

Aquaporins are water-selective channel proteins, which are constituents of plant membranes and have been shown to be salt and drought inducible (Kjellbom et al., 1999). Although aquaporins constitute a very large multigene family, *PIP1* antisense reduction of the mRNA levels in *Arabidopsis* demonstrated the importance of these molecules in plant water transport (Kaldenhoff et al., 1998). Interestingly, the plants were able to compensate for the reduced cellular water permeability in the transgenic plants by increasing the size of the root system by five fold, supporting the importance of root mass in plant adaptation to water deficit.

Late embryogenesis abundant (LEA) proteins (Dure, 1992), also known as dehydrins (Close, 1997), accumulate in response to the plant hormone abscissic acid (ABA) and environmental stress conditions that involve dehydration, such as drought salinity and cold. While the function of these and novel hydrophilic proteins induced by cold adaptation still remains unclear, they have been postulated to interact with hydrophobic surfaces through their amphipatic α -helices and thus protect macromolecules and membranes under conditions of decreasing water activity and during seed desiccation. The protective effect of the dehydrin gene family on dehydration stress tolerance has been tested to a limited extent. High levels of constitutive accumulation were obtained in transgenic rice from the barley *HVA1* gene, a member of the LEA III group of proteins (Xu et al., 1996). The *HVA1* expression in both leaves and roots from the rice actin 1 gene promoter provided significantly increased tolerance to salinity and water

deficit, however constitutive expression required metabolic channeling with potential yield cost and also led to frequent gene silencing in this system (Cheng et al., 1998). Subsequent efforts have aimed to engineer promoter elements that would be stress inducible, by modifying the rice actin 1 promoter with the ABA response (ABRC1) element from the barley HVA1 gene, in order to make future transgene expression in rice stress-inducible (Su et al., 1998).

The hydrophilic, boiling resistant *COR* genes in *Arabidopsis* are grouped in four gene families, with at least one member of each family being induced by low temperature, drought, high salinity and ABA. Constitutive expression of the single *COR15a* gene product, targeted to chloroplast stroma, showed a positive effect on freezing protoplast tolerance for a very limited temperature range (Artus et al., 1996). Subsequent experiments, in which a transcriptional activator was introduced in transgenic plants to induce additional *Arabidopsis COR* genes have been more successful in providing freezing tolerance as measured by electrolyte leakage (Jaglo-Ottosen et al., 1998) and will be discussed in more detail below. Spinach *CAP* (cold acclimation protein) genes (Guy and Haskell, 1987) are also induced by cold, dehydration and ABA and encode boiling soluble polypeptides. Transgenic tobacco constitutively expressing two of these genes had a somewhat slowed rate of cellular damage by freezing (Kaye et al., 1998), suggesting an effect on membrane stability during freezing, but as in the case of single *COR* gene introduction, the overall improvement in tolerance was small.

1.3.3. Overexpression of gene products to counteract oxidative injury and photoinhibition during abiotic stress.

Photoinhibition of photosynthesis accompanies plant salt, drought and cold stress, especially at high light intensities. Oxidative stress-sensitive *Arabidopsis* mutants deficient in ascorbic acid have shown the essential nature of detoxification for environmental stress resistance (Conklin et al., 1996). Studies on molecular mechanisms to prevent photoinhibition, damage to photosystem II (PSII) and cellular membranes by reactive oxygen species (ROS) have focused on overproduction of enzymes with potential to eliminate ROS, with the expectation that the transgenic plants would be more resistant to oxidative and other abiotic stress conditions (Allen, 1995; Foyer et al., 1994; Foyer et al., 1998).

Overexpression of superoxide dismutase (SOD) in tobacco was shown to protect plants from oxidative stress as induced by high light and low temperature, while maintaining better photosynthetic rates under the stress conditions (Gupta et al., 1993). Mn-SOD reduced cellular damage from oxidative stress in transgenic tobacco (Bowler et al., 1991). Transgenic alfalfa overexpressing Mn-SOD from tobacco showed enhanced tolerance to freezing stress (McKersie et al., 1993) and reduced injury from water deficit, while maintaining the requisite field performance (McKersie et al., 1996). To date, this remains the only study in which field performance of the transgenic crop has been tested. However, McKersie and co-workers (McKersie et al., 1999) also caution that pre-screening of the transgenics in the greenhouse was not effective in predicting individual transgenic plants with improved field performance and winter survival. Overproduction of Fe-SOD in tobacco chloroplasts also provided improved oxidative stress resistance (Van Camp et al., 1996). Subsequent work with poplar, overexpressing chloroplast-targeted Fe-SOD suggested that increased levels of the enzyme can protect PSII from overreduction at low CO₂ concentrations, but could not protect PSII efficiency under these conditions (Arisi et al., 1998). Fe-SOD overproduction also did not prevent photoinhibition of PSII in the poplar (Tyystjarvi et al., 1999), indicating that ROS scavenging enzymes may have indirect effects in the chloroplast.

Increased protection to salt and drought stress has also been reported by overexpression of the cytosolic ascorbate peroxidase (APX), but no protection was obtained with the chloroplast targeted isoform (Torsethaugen et al., 1997). Glutathione-S-transferase overexpression in tobacco provided protection against both cold and salt stress in seedlings (Roxas et al., 1997). In contrast, glutathione reductase overproduction in the cytosol showed somewhat reduced photoinhibition in transgenic tobacco, but did not enhance repair of damage caused by the photoinhibition (Tyystjarvi et al., 1999). Glutathione homeostasis has also been altered by overexpression of *Brassica* glyoxalase I in tobacco (Veena et al., 1999).

While cellular compartmentalization of the different antioxidant enzymes and the interacting pathways contribute to the complexity of interpreting these results, recent studies also suggest that H₂O₂ is likely to play a very important role in the abiotic stress signaling pathway (Kovtun et al., 2000b) and connect oxidative stress, auxin and cell cycle regulation through the mitogen-activated kinase pathway in plants (Hirt, 2000). This connection

may explain some observed cross tolerance to different stress reactions with the resulting broad based activation of endogenous genes. Such has been exhibited in *Arabidopsis* expressing the bacterial choline oxidase gene for glycinebetaine synthesis, where the by-product was increased levels of H_2O_2 , which might have led to the observed stimulation of antioxidant scavenging enzymes (Alia et al., 1999) and provided protection to salt, cold, heat and light stress (Table 1). More experiments will be required to provide us with understanding of the mechanistic balance required in oxidative stress damage versus its role as a signal in gene activation against stress.

1.4. TRANSGENIC APPROACH TO INCREASE RESISTANCE TO OSMOTIC STRESS BY MULTIPLE GENE REGULATION THROUGH SIGNALING AND TRANSCRIPTION

Specific and coordinate mRNA accumulation in response to salt, drought and cold stress in plants has been reviewed (Bray, 1997; Ingram and Bartels, 1996; Shinozaki and Yamaguchi-Shinozaki, 1999; Thomashow, 1999; Winicov, 1998). The signaling pathways and molecular mechanisms responsible for this coordinate transcript accumulation have been also reviewed recently (Hasegawa et al., 2000; Shinozaki et al., 1999). Most current models for stress gene activation through these pathways are built predominantly by analogy to the yeast osmosensing and signaling pathways, since many of the signaling gene products identified in plants can complement mutations in the yeast system. However, it is also becoming apparent that the diversity and number of functional groups of genes activated during the stress response is likely to contain a temporal and selective component (Winicov, 2000b), namely: 1) some of the induced genes may represent a temporal adjustment to metabolic imbalance caused by the stress; 2) some may initiate a cascade of responses, but are not necessary to maintain subsequent change, suggesting perhaps another cascade required to return to the unstressed metabolic state; and 3) some of the activated gene products contribute directly to the cellular alterations that enable tolerance, as demonstrated by survival and productive growth.

Native differences in resistance to salt, drought and cold/freezing stress that exist between crop plant varieties indicate that different levels of utilization of similar genetic information can have a positive effect on survival and yield under stress conditions. Our improving understanding of the physiological relatedness of gene products in emerging pathways, has suggested that it

might be possible to approach the problem of improvements in tolerance by manipulating: 1) selected signaling systems that lead to enhanced expression of genes that counter the stress response, or 2) those transcription factors that are limiting in expression of genes that can provide tolerance to osmotic stress. Data from several laboratories indicate that this may indeed be a feasible approach by which to regulate groups of related genes constitutively or in a stress inducible manner and we can expect more progress in this area in the coming years.

1.4.1. Overexpression of stress signal components to improve resistance to osmotic stress.

An increasing number of kinases and phosphatases have been identified that respond to salt, drought and cold stress. However, the pathways themselves have remained largely unresolved because of the large number of genes encoded by these gene families and the uncertainty about their tissue and developmental specificity. The pathways include both calcium dependent protein kinases (CDPK) and the mitogen-activated protein kinase cascade (Harmon et al., 2000; Mizoguchi et al., 2000). Many of the putative signaling components are induced by salt, drought, cold, ABA, light, plant hormones and even pathogen infection (Moller and Chua, 1999), although there is some stress specificity for induction of some protein kinases. For example the Ca^{+2} inducible kinases ATCDPK1 and ATCDPK2 are rapidly induced by salt and drought, but are not induced by cold, heat or ABA (Urao et al., 1994) and the mitogen-activated kinase MMK4 is induced by drought and cold, but not salt and ABA (Jonak et al., 1996), indicating that it might be possible to distinguish between the intersections of the different stress pathways. However, the complexity of the interacting signaling pathways is further underscored by experiments that show improved salt tolerance in plants constitutively overexpressing either a protein phosphatase or kinase as discussed below.

Calcium dependent protein kinases have been shown to activate a stress-inducible promoter responding to cold, salt dark and ABA (Sheen, 1996), indicating a broad range of activities for these proteins in stress signaling. A more dynamic mode of action and distinct Ca^{+2} signaling pathways appear to regulate calmodulin expression by different stress conditions (van der Luit et al., 1999). Calcium signaling through a calcineurin-like pathway in salt stress

response also has been implied by the salt hypersensitivity of the SOS3 mutation in *Arabidopsis* (Liu and Zhu, 1998). The relevance of these signaling pathways to improving salt-tolerance was tested in tobacco by introducing a modified form of yeast calcineurin (CaN) to express a constitutively activated Ca^{+2} -and calmodulin-dependent protein phosphatase (Pardo et al., 1998). The ubiquitous CaN expression in transgenic tobacco was associated with improved resistance to salt stress, and grafting experiments indicated that roots provided primarily those functions significant for the improved tolerance, but the specificity and selectivity of CaN interactions with the endogenous signaling cascades remains to be demonstrated for this system.

Recently it has been shown that manipulation of mitogen-activated protein kinases that mediate signal transduction of growth control, cell cycle and stress responses can also enhance tolerance to abiotic stress (Kovtun et al., 2000a). The *Arabidopsis* and tobacco MAPKKs, ANP1 and NPK1 respectively, mediate H_2O_2 signal transduction in oxidative stress. The constitutively expressed NPK1 was shown to function in oxidative stress signaling, and it had a negative effect on auxin induced gene expression, but provided enhanced tolerance to abiotic stress without a notable effect on growth and development. Interestingly it also did not activate some previously described drought, cold and ABA signaling pathways, implying that it may function in selective cascades. These results emphasize the need for better understanding of the connectivity and specific targeting of the many kinases and phosphatases and also support the possibility of achieving tolerance by different pathways. While it is recognized that hormones, such as ABA (Leung and Giraudat, 1998), ethylene (Kieber et al., 1993) and auxin (Qin and Zeevaart, 1999) may have a significant role in abiotic stress responses, the limiting steps in their metabolic activation and signal transduction pathways are not yet sufficiently clear to provide predictable options for their manipulation in presence of the extensive crosstalk in plant signaling (Genoud and Metraux, 1999).

1.4.2. Overexpression of transcription factors to activate multiple genes and provide protection to osmotic stress.

Relatively few transcription factors have been identified to date that bind to promoter elements in genes regulated by salt, drought and cold stress in

plants and thus could potentially provide a means for regulating multiple genes associated with the stress response. An early candidate was the plant bZIP protein EmBP-1 (Guiltinan et al., 1990) that recognizes the ABA response element present in a number of genes inducible by abiotic stress and ABA application. Production of the EmBP-1 protein is enhanced during embryo maturation and in other tissues during salt/drought stress. However, overexpression of a truncated dominant negative version of the EmBP-1 protein in tobacco showed no improvement in stress tolerance, but instead revealed an important developmental function for this protein in vegetative tissue (Eckardt et al., 1998). Transient transactivation experiments have shown that both *myc* and *myb* proteins function as transcriptional activators of the dehydration and ABA inducible gene *rd22* (Abe et al., 1997), but the effects of these transcription factors on general plant growth and tolerance to osmotic stress has not been reported. New information will be necessary for the mechanistic understanding of ABA mediated activation of gene expression during abiotic stress in vegetative tissues.

Table 2 lists transcription factors currently correlated with responses to salt, drought and cold and results from their overexpression, where known. Dehydration induced expression of two homeodomain-leucine zipper proteins in *Craterostigma plantagineum* was demonstrated (Frank et al., 1998), but their function has not yet been assessed in transgenic plants. The TFIIIA-type zinc-finger factor STZ was shown to increase in *Arabidopsis* after salt treatment and was able to activate transcription and complement the salt-sensitive phenotype in yeast (Lippuner et al., 1996), but its role for salt tolerance in plants remains to be tested. The more extensive studies with yeast mutants have identified a significant number of transcription factors that differentially regulate gene expression by short-term and long-term osmotic stress and indicate the likelihood of a multi-tiered system in transcriptional regulation with additional specificity conferred through different signaling pathways (Matheos et al., 1997; Rep et al., 1999).

TABLE 2. Transcription factors associated with osmotic stress responses with potential for other gene regulation to increase stress tolerance.

Factor	Type	Stress	Comments
(Reference)			
CPM7	<i>myb</i>	Dehydration	Not tested in transgenics (Iturriaga et al., 1996)
ATMYB2	<i>myb</i>	Dehydration	Transcriptional activator in dehydration and ABA-induction of <i>rd22</i> expression (Abe et al., 1997), not tested for tolerance.
rd22BP1	<i>myc</i>	Dehydration	Transcriptional activator in dehydration and ABA-induction of <i>rd22</i> expression (Abe et al., 1997), not tested for tolerance.
CPHB-1	Homeodomain-leucine zipper	Dehydration	Not tested in transgenics (Frank et al., 1998)
CPHB-2	Homeodomain-leucine zipper	Dehydration ABA	Not tested in transgenics (Frank et al., 1998)
STZ	Cys ₂ His ₂ Zn-finger	Salt	Complements salt sensitive phenotype in yeast, not tested in plants (Lippuner et al., 1996)
ZPT2-2	Cys ₂ His ₂ Zn-Finger	Dehydration Cold, UV-B	Also induced by wounding, jasmonic acid hormonal responses in leaves (van der Krol et al., 1999), not tested for stress function in transgenics
CBF1	AP2/ERBP	Cold	Binds CRT/DRE element TACCGACAT, constitutive overexpression induces COR gene expression and improves freezing acclimation in plants (Jaglo-Ottosen et al., 1998)
DRB1A	AP2/ERBP	Cold	Binds CRT/DRE element TACCGACAT, constitutive overexpression induces <i>rd29A</i> and other genes with DRE element, dwarf phenotype, freezing and dehydration tolerance (Liu et al., 1998). Stress inducible overexpression eliminates dwarf phenotype (Kasuga et al., 1999)
DRB2A	AP2/ERBP	Dehydration	Binds CRT/DRE element TACCGACAT, constitutive overexpression induces <i>rd29A</i> weakly, dwarf phenotype, freezing and dehydration tolerance (Liu et al., 1998).
Alfin1	Cys ₄ -His/Cys ₃ Zn-finger	Salt	Binds GTGGNG or GNGGTG, constitutive overexpression induces <i>MsPRP2</i> in roots, increases salt tolerance (Winicov and Bastola, 1999) and root growth (Winicov, 2000a)

Identification of the CRT/DRE (drought responsive element) in many genes regulated by cold, drought and salt through an ABA independent pathway

proteins: 1) DRB1A, CBF1= DRB1B, DRB1C and 2) DRB2A and DRB2B, that bind to the 9-bp conserved sequence element (Shinwari et al., 1998; Stockinger et al., 1997; Liu et al., 1998). Constitutive overexpression of the CBF1 in transgenic *Arabidopsis* resulted in increased transcription of the COR genes without exposure to cold and led to increased freezing tolerance as measured by electrolyte leakage (Jaglo-Ottosen et al., 1998) and activation of multiple biochemical changes (Gilmour et al., 2000) thus showing CBF1 to be an important regulator in the cold-acclimation process. Similarly, constitutive overexpression of DREB1A and DREB2A induced enhanced expression of target genes under non-stressed conditions in the transgenic plants and provided some freezing and dehydration tolerance, but unlike the CBF1 containing transgenics, showed severe growth retardation under normal conditions (Liu et al., 1998). Subsequent experiments with transgenic overexpression of DREB1A from the stress inducible *rd29A* promoter allowed for normal growth under control conditions while providing the transgenic plants with enhanced levels of DREB1A under stress conditions and improved stress resistance to freezing, drought and salinity as measured by plant survival (Kasuga et al., 1999).

Recently, we have shown, that the putative transcription factor *Alfin1*, initially cloned by differential screen from a cDNA library constructed with mRNA from salt tolerant alfalfa callus (Winicov, 1993), does bind DNA in a sequence specific manner and recognizes these DNA elements in the promoter of the alfalfa *MsPRP2* gene (Bastola et al., 1998). Interestingly, both *Alfin1* and *MsPRP2* are salt inducible and are expressed predominantly in roots. Transgenic plants overexpressing *Alfin1* from the CaMV 35S promoter in both roots and shoots were normal in appearance, showed improved salt tolerance and enhanced levels of *MsPRP2* mRNA in the roots, consistent with *Alfin1* protein acting as a transcriptional regulator (Winicov and Bastola, 1999). Since *Alfin1* antisense expressing transgenics could not regenerate roots that could survive in soil and *Alfin1* overexpressing plants showed two to four fold enhanced root growth when compared to the untransformed parent plants, it is likely that *Alfin1* is an essential root transcription factor that modulates the expression of genes in the root (Winicov, 2000a) and by doing so enhances root growth in a manner that in turn improves resistance to salinity stress. Since *Alfin1* belongs to a small and plant specific transcription factor family (Reichman et al., 2000) and is conserved between alfalfa, rice and *Arabidopsis* (Winicov and Bastola,

1997), it could be useful in improving tolerance and root growth in other crop plants.

The results obtained from different laboratories, which show single-transcription factor overexpression providing increased tolerance to osmotic stress indicate that it is feasible to manipulate the expression of multiple genes by transcriptional regulators.

2. HEAT STRESS MANAGEMENT

Heat stress in plants as in all eukaryotes shuts down or decreases most gene expression, and at sublethal doses induces specific heat shock (HS) genes that allow the plant to survive and recover from the elevated temperatures (Vierling, 1991). This is accomplished by heat shock proteins (HSPs) that act as molecular chaperones preventing aggregation and enabling ATP dependent refolding of denatured proteins; and another group of proteins, which includes ubiquitin and proteases, that target the denatured proteins for degradation. Transcriptional activation of the many HSPs occurs in response to signal transduction cascade which activates transcription of different members of the heat stress transcription factor (HSF) gene families (Nover et al., 1996). The HSF proteins in plants activate stress genes and are themselves regulated by signals from heat stress, cell cycle and developmental cues, as well as the HSPs themselves (Schoffl et al., 1998).

Inducible thermotolerance provides plants with a well characterized and effective system with which to counteract short term increases in temperature encountered seasonally during diurnal high temperature cycle. However, adaptation or preconditioning appears to be essential for heat tolerance as has been shown for cold acclimation discussed above.

2.1. HEAT SHOCK PROTEINS-DIVERSE ROLES IN DEVELOPMENT AND ABIOTIC STRESS

Plant HSPs belong to five structurally distinct classes. The chaperonins, or HSP100, HSP90, HSP70 and HSP60 and the small HSPs, which are the most abundant in plants. The small HSPs (monomeric 15 to 30 kDa) consist of five conserved gene families, encoding proteins found in cytosol (class I and II), mitochondria, chloroplasts and the endoplasmic reticulum (Waters et

al., 1996).

Although the HSPs were first identified as proteins induced by heat stress, some of the HSPs are expressed constitutively, some are under cell cycle or developmental control and some may provide protection in desiccation tolerance of seeds (Wehmeyer and Vierling, 2000) as indicated by dual regulation of an HSP promoter during embryogenesis (Carranco et al., 1999). Developmental regulation of a subset of HSPs suggests that these gene products may have specific functions that differ from those utilized during environmental stress. Interestingly, HSPs (Athsp70-1, Athsp81-2 and ubiquitin extension protein) also have been identified as genes induced in ABA independent early-response to dehydration (Kiyosue et al., 1994) and the chloroplast small HSP may be associated with protection from oxidative stress (Harndahl et al., 1999). Small HSP mRNAs have been found to accumulate under a variety of developmental and stress conditions, besides heat stress (Waters et al., 1996). Rice HSP110 was found to be highly inducible in 5 day seedlings by salinity, drought and cold stress (Singla et al., 1997), indicating an expanded role for HSPs at different stages of development in stress responses.

Specific functions of individual HSPs are still being clarified, but the current model suggests that small HSPs prevent thermal aggregation of proteins by binding to the denatured forms (Lee et al., 1997). Once bound, the protein becomes a substrate for refolding by large protein complexes containing HSC70/HSP70 (Lee and Vierling, 2000). The HSP100 class present in plants, yeast and bacteria, but not *Drosophila*, promotes reactivation of aggregated proteins (Parsell and Lindquist, 1993). Plant 101 kDA was able to complement a yeast mutation and improve acquired thermotolerance, thus indicating a role for this protein in recovery from heat stress (Lee et al., 1994; Schirmer et al., 1994). Recent characterization of the *hot1* mutant in *Arabidopsis*, which is unable to acquire thermotolerance, as point mutation in the HSP101 protein ATP-binding domain provides strong evidence that this protein is essential for thermotolerance in plants (Hong and Vierling, 2000). The *hot1* seeds showed greatly reduced basal thermotolerance, also supporting the role for HSP101 accumulation in seed function. Since the same study also identified three additional loci important for acquired thermotolerance, it is apparent that HSP101 is not solely responsible for thermotolerance in plants but is part of a multigene response.

2.2. TRANSGENIC APPROACH TO INCREASE THERMOTOLERANCE IN PLANTS

Most strategies for improving heat stress tolerance in plants have focused on altering the constitutive levels of individual HSPs, or groups of HSPs by transgenic overexpression of HSFs as reviewed (Gurley, 2000) and summarized in Table 3, although it is interesting to see that improved thermotolerance has also been achieved with manipulation of membrane components and even engineering synthesis of the osmolyte glycinebetaine. Unlike osmotic stress, the HS response is temporary, lasting only about 2 hours after stress, although the HSPs show a half life of about 24 hr, suggesting tight regulation of the stress response

TABLE 3. Transgenes with effects on increasing thermotolerance in plants.

Gene	Function	Reference
<i>Athsf1</i> -GUS	HSF (constitutive expression of HSP18)	(Lee et al., 1995)
<i>Athsf3</i> -GUS <i>Athsf3</i>	HSF (constitutive expression of class I small HSPs)	(Prandl et al., 1998)
<i>HSP101</i>	HSP (HSP101 constitutive expression)	(Queitsch et al., 2000)
<i>codA</i>	Choline oxidase, engineered glycine betaine synthesis and accumulation	(Alia et al., 1998)
<i>FAD7</i>	Silencing chloroplast ω -3 desaturase, lower level of trienoic acids	(Murakami et al., 2000)
<i>fad7fad8</i>	mutant in two chloroplast ω -3 desaturase genes, decreased trienoic acids	(Murakami et al., 2000)

Initial ectopic expression of an HSF1-glucuronidase (GUS) or GUS-HSF1 fusion protein in transgenic *Arabidopsis* led to constitutive expression at normal temperatures of small HSP18 that were about 20% of those observed after heat stress. This increase correlated with an increased level of basic, but no induced thermotolerance (Lee et al., 1995). Similar experiments with *Arabidopsis* HSF3- and HSF4-GUS fusion proteins in transgenic plants yielded increased HSP levels at normal temperatures correlated only with expression of HSF3 fusion protein (Prandl et al., 1998) and increased basal

thermotolerance by 2°C in the transgenic plants, but had no effect on acquired thermotolerance. The transgenic expression of HSF-GUS and low levels of constitutive expression of HSPs had no effect on the phenotype in absence of heat shock.

Direct manipulation of HSP101 levels in transgenic *Arabidopsis* by constitutive expression of HSP101 in sense or antisense orientation has demonstrated that HSP101 levels are crucial for both basal and acquired thermotolerance (Queitsch et al., 2000). Plants with reduced HSP101 did not survive at high temperatures with preconditioning that permitted wild type survival. Constitutively overexpressing plants with 40 to 85% levels of HSP101 induced by pretreatment, were able to survive and recover from a shorter heat shock at 45°C without pretreatment. These results indicate that other factors are needed in addition to HSP101 to provide the full protection obtained from preconditioning, but HSP101 alone provided a significant improvement in thermotolerance of both plants and seedlings. Since HSP101 depleted seeds from transgenic plants expressing antisense HSP101, or plants exhibiting cosuppression of HSP101, were hypersensitive to heat shock compared to wild type plants, this also demonstrates the importance of HSP101 in seed resistance to the effect of severe heat stress during germination. Since transgenic plants overexpressing HSP101 showed no detrimental effects on normal growth and development it should be possible to reduce the dependence of other plants to preconditioning for improvement of thermotolerance.

Although most of the effort to date in manipulation of thermotolerance has been directed against HSPs, another strategy has been aimed at altering membrane composition by genetic engineering in order to increase thermotolerance (Murakami et al., 2000). Chloroplast trienoic fatty acid synthetase gene expression was inhibited in tobacco through gene silencing by introduction of the chloroplast-localized ω -3 desaturase gene (*FAD7*). This resulted in reduced levels of chloroplast trienoic acids in several transgenic lines with positive correlation to maintenance of photosynthesis at elevated temperatures and general improvement in thermotolerance of the transgenic plants. Similar results were obtained with *Arabidopsis fad7fad8* mutant, lacking two chloroplast-localized ω -3 fatty acid desaturases (McConn et al., 1994) indicating that the lower levels of trienoic fatty acids in chloroplast membranes were protective in surviving heat stress. No differences in growth were observed for the transgenic and wild type tobacco

between 15 and 25°C, but similar growth inhibition was observed for all plants below 10°C. This is somewhat puzzling since slightly elevated trienoic fatty acid level due to overexpression of the *FAD7* ω -3 desaturase gene (Kodama et al., 1994) had a protective effect against chilling damage in tobacco seedlings.

An example of cross-tolerance against different abiotic stress regimes was observed in *Arabidopsis* transformed with the *codA* gene for choline oxidase to engineer high levels of glycinebetaine as an osmoprotectant for drought, cold and improved salinity tolerance. The accumulation of glycinebetaine in seeds correlated with enhanced tolerance to high temperatures during germination and growth of seedlings (Alia et al., 1998). Associated decreases of HSP70 homologue induction in the transgenics at high temperature was interpreted as osmoprotectant alleviated high temperature stress.

CONCLUSIONS AND FUTURE PROSPECTS

Plants manifest a complex and dynamic range of molecular reactions when exposed to abiotic stress, connected through a web of signal cascades with intersecting and sometimes additive effects. Oxidative reactions appear to be a common byproduct of cold, heat, drought and salinity stress and the role of calcium seems equally ubiquitous as a participant in the various signaling systems activated by these stress conditions (Bowler and Fluhr, 2000). The impact of these signals on gene expression and metabolism is multifaceted and will require our understanding of those gene functions that are optimal for tolerance as differentiated from those that are marginally helpful, since it has been amply demonstrated to date, that marginal improvements in tolerance can be achieved by manipulation of many different genes.

We have learned from the current studies on abiotic stress responses in plants that acute stress responses may or may not be sustainable constitutively without penalty on yield. On the other hand, adaptation or acclimation plays a significant role in most abiotic stress resistance and as clearly shown by results in expression of genes important for cold and heat adaptation, the incremental benefits of constitutive expression may be substantial for survival. For these responses an inducible promoter may not offer additional advantages unless a significant improvement in the threshold of survival temperature could be obtained. Because of the multigenic character of the stress response, the question of additivity of the protective effects of different

gene products also needs to be explored.

The emphasis to date has been in utilization of our knowledge of gene expression in response to short term abiotic stress. Diurnal temperature changes are temporal and the molecular responses reflect the duration of the natural stress conditions, which support a major role for pre-conditioning in improving tolerance or recovery from the stress. Seasonal temperature changes are more long term, with more gradual adaptation and may or may not involve additional changes in gene expression. For engineering tolerance to drought and salinity the scope of stress resistance becomes generally long term and induction of different metabolic systems, normally stress limited, may provide incremental improved tolerance. Presently we have little information on long term tolerance from these physiological systems which may have a more significant component of yield, another polygenic trait. Associated with yield may be increased stress resistance. Recent interesting examples of this can be seen in transgenic potato and alfalfa. Phytochrome B overexpression in potato led to higher photosynthetic performance, resistance to prolonged light stress (oxidative stress) and also increased tuber yield, presumably due to delayed senescence of the leaves (Thiele et al., 1999). Transgenic alfalfa overexpressing the transcription factor *Alfin1*, showed markedly improved root growth with associated improvement in salinity tolerance (Winicov, 2000a; Winicov and Bastola, 1999).

The polygenic trait of differential tolerance to abiotic stress raises the question if different metabolic systems become limiting under stress in different plants. This raises the specter of tailoring tolerance and markers for individual crops with its inherent complications. In order to manipulate functionally related genes with a minimum of undesirable side effects it will be necessary to understand those factors that are limiting in their expression as close to the end product in the signaling or transcription steps as possible, because of the inter-relatedness of signal and metabolic pathways leading to undesirable side effects. Successful identification of genes able to accomplish this would provide reliable molecular markers for breeding of crop plants. Alternatively, ingenious application of selective systems in biology has been fruitful in the past for molecular manipulations and should be considered seriously in future for attaining the goal of more tolerant crops.

Genetic transformation of plants with a variety of genes involved in the

abiotic stress responses has demonstrated that incremental improvements in stress resistance can be obtained, in many cases with minimal pleotropic effects on the plant. Recent examples in manipulation of signaling pathways and transcription factors for gene regulation that counteract the abiotic stress responses will continue to be expanded in order to channel efforts toward polygenic control of resistance to abiotic stress.

ACKNOWLEDGMENTS

Because of space constraints, not all references could be included and my apologies in this regard. Many thanks to colleagues for discussions and making available unpublished data.

REFERENCES

- Abe, H., Yamaguchi-Shinozaki, K., Urao, T., Iwasaki, T., Hosokawa, D., and Shinozaki, K. (1997). Role of *Arabidopsis* MYC and MYB homologs in drought- and abscisic acid-regulated gene expression. *Plant Cell* **9**, 1859-1868.
- Alia, Hayashi, H., Sakamoto, A., and Murata, N. (1998). Enhancement of the tolerance of *Arabidopsis* to high temperatures by genetic engineering of the synthesis of glycinebetaine. *Plant J.* **16**, 155-161.
- Alia, Kondo, Y., Sakamoto, A., Nonaka, S., Hayashi, H., Saradhi, P. P., Chen, T. H. H., and Murata, N. (1999). Enhanced tolerance to light stress of transgenic *Arabidopsis* plants that express the *codA* gene for a bacterial choline oxidase. *Plant Mol. Biol.* **40**, 279-288.
- Allen, R. D. (1995). Dissection of oxidative stress tolerance using transgenic plants. *Plant Physiol.* **107**, 1049-1054.
- Apse, M. P., Aharon, G. S., Snedden, W. A., and Blumwald, E. (1999). Salt tolerance conferred by overexpression of a vacuolar Na^+/H^+ antiport in *Arabidopsis*. *Science* **285**, 1256-1258.
- Arisi, A.-C. M., Cornic, G., Jouanin, L., and Foyer, C. H. (1998). Overexpression of iron superoxide dismutase in transformed poplar modifies the regulation of photosynthesis at low CO_2 partial pressures or following exposure to the prooxidant herbicide methyl viologen. *Plant Physiol.* **117**, 565-574.
- Artus, N. N., Uemura, M., Steponkus, P. L., Gilmour, S. J., Lin, C. T., and Thomashow, M. F. (1996). Constitutive expression of the cold-regulated *Arabidopsis thaliana* *COR15a* gene affects both chloroplast and protoplast freezing tolerance. *Proc. Natl. Acad. Sci. U.S.A.* **93**, 13404-13409.
- Bastola, D. R., Pethe, V. V., and Winicov, I. (1998). Alfin1, a novel zinc-finger protein in alfalfa roots that binds to promoter elements in the salt-inducible *MsPRP2* gene. *Plant Mol. Biol.* **38**, 1123-1135.
- Bohnert, H. J., and Sheveleva, E. (1998). Plant stress adaptations-making metabolism move. *Curr. Opinion in Plant Biol.* **1**, 267-274.
- Bowler, C., and Fluhr, R. (2000). The role of calcium and activated oxygens as signals for

- controlling cross-tolerance. *Trends in Plant Sci.* **5**, 241-246.
- Bowler, C., Slooten, L., Vandenbranden, S., Rycke, R. D., Botterman, J., Sybesma, C., Montagu, M. V., and Inze, D. (1991). Manganese superoxide dismutase can reduce cellular damage mediated by oxygen radicals in transgenic plants. *EMBO* **10**, 1723-1732.
- Bray, E. A. (1997). Plant responses to water deficit. *Trends in Plant Sci.* **2**, 48-54.
- Carranco, R., Almoguera, C., and Jordano, J. (1999). An imperfect heat shock element and different upstream sequences are required for the seed-specific expression of a small heat shock protein gene. *Plant Physiol.* **121**, 723-730.
- Chapman, K. D. (1998). Phospholipase activity during plant growth and development and in response to environmental stress. *Trends in Plant Sci.* **3**, 419-426.
- Cheng, W., Su, J., Zhu, B., Jayaprakash, T. L., and Wu, R. (1998). Development of transgenic cereal crop plants that are tolerant to high salt, drought and low temperature. In "Frontiers in Biology: The Challenges of Biodiversity." (C. H. Chou and K. T. Shao, Eds.), pp. 115-122. Academia Sinica, Taipei.
- Close, T. (1997). Dehydrins: A commonality in the response of plants to dehydration and low temperature. *Physiol. Plant.* **100**, 291-296.
- Conklin, P. L., Williams, E. H., and Last, R. T. (1996). Environmental stress sensitivity of an ascorbic acid-deficient Arabidopsis mutant. *Proceedings of the National Academy of Sciences of the USA* **93**, 9970-9974.
- Deutch, C. E., and Winicov, I. (1995). Post-transcriptional regulation of a salt-inducible alfalfa gene encoding a putative chimeric proline-rich cell wall protein. *Plant Molecular Biology* **27**, 411-418.
- Dure, L. I. (1992). The LEA proteins of higher plants. In "Control of Plant Gene Expression" (D. P. S. Verma, Ed.), pp. 325-335. CRC Press, Boca Raton, FL.
- Eckardt, A. N., McHenry, L., and Guiltinan, M. J. (1998). Overexpression of EmBP, a dominant negative inhibitor of G-box-dependent transactivation, alters vegetative development in transgenic tobacco. *Plant Mol. Biol.* **27**, 411-418.
- Flowers, T. J., Koyama, M. L., Flowers, S. A., Sudhakar, C., Singh, K. P., and Yeo, A. R. (2000). QTL: their place in engineering tolerance of rice to salinity. *J. Exp. Botany* **51**, 99-106.
- Foolad, M. R., and Jones, R. A. (1993). Mapping salt-tolerance genes in tomato (*Lycopersicon esculentum*) using trait-based marker analysis. *Theor. Appl. Genet.* **87**, 184-192.
- Foyer, C. H., Descourvieres, P., and Kunert, K. J. (1994). Protection against oxygen radicals: an important defence mechanism studied in transgenic plants. *Plant Cell Envir.* **17**, 507-523.
- Foyer, C. H., Kingston-Smith, A. H., Harvinson, J., Arisis, A.-C. M., Jouanin, L., and Noctor, G. (1998). The use of transformed plants in the assesment of physiological stress responses. In "Responses of plant metabolism to air polution and global change." (L. J. De Kok and I. Stulen, Eds.), pp. 251-261. Backhuys, Leiden, The Netherlands.
- Frank, W., Munnik, T., Kerkmann, K., Salamini, F., and Bartels, D. (2000). Water deficit triggers phospholipase D activity in the resurrection plant *Craterostigma plantagineum*. *Plant Cell* **12**, 111-123.
- Frank, W., Phillips, J., Salamini, F., and Bartels, D. (1998). Two dehydration-inducible transcripts from the resurrection plant *Craterostigma plantagineum* encode interacting homeodomain-leucine zipper proteins. *Plant J.* **15**(3), 413-421.

- Gage, D. A., Rhodes, D., Nolte, K. D., Hicks, W. A., Leustek, T., Cooper, A. J. L., and Hanson, A. D. (1997). A new route for synthesis of dimethyl-sulphoniopropionate in marine algae. *Nature* **387**, 891-894.
- Galiba, G., Quarrie, S. A., Sutka, J., Moroonov, A., and Snape, J. W. (1995). RFLP mapping of the vernalization (*Vrn1*) and frost resistance (*Fr1*) genes on chromosome 5A of wheat. *Theor. Appl. Genet.* **90**, 1174-1179.
- Galiba, G., Simon-Sarkadi, Kocsy, L., Salgo, G., and Sutka, A. (1992). Possible chromosomal location of genes determining the osmoregulation of wheat. *Theor. Appl. Genet.* **85**, 415-418.
- Gilmour, S. J., Sebolt, A.M., Salazar, M.P., Everard, J.D., and Thomashow, M.F. (2000). Overexpression of the Arabidopsis CBF3 transcriptional activator mimics multiple biochemical changes associated with cold acclimation. *Plant Physiol.* **124**, 1854-1865.
- Genoud, T., and Metraux, J.-P. (1999). Crosstalk in plant cell signaling: structure and function of the genetic network. *Trends in Plant Sci.* **4**, 503-507.
- Guiltinan, M. J., Marcotte, W. R., and Quatrano, R. S. (1990). A plant leucine zipper protein recognizes an abscisic acid response element. *Science* **250**, 267-270.
- Gupta, A. S., Webb, R. P., Holaday, A. S., and Allen, R. D. (1993). Overexpression of superoxide dismutase protects plants from oxidative stress. *Plant Physiol.* **103**, 1067-1073.
- Gurley, W. B. (2000). HSP101: a key component for the acquisition of thermotolerance in plants. *Plant Cell* **12**, 457-460.
- Guy, C. L., and Haskell, D. (1987). Induction of freezing tolerance in spinach is associated with the synthesis of cold acclimation induced proteins. *Plant Physiol.* **84**, 872-878.
- Halfiter, U., Ishitani, M., and Zhu, J.-K. (2000). The Arabidopsis SOS2 protein kinase physically interacts with and is activated by the calcium-binding protein SOS3. *Proc. Natl. Acad. Sci. U.S.A.* **97**, 3735-3740.
- Harmon, A. C., Gribskov, M., and Harper, J. F. (2000). CDPKs-a kinase for every Ca⁺² signal. *Trends in Plant Sci.* **5**, 154-159.
- Härndahl, U., Buffoni Hall, R., Osteryoung, K. W., Vierling, E., Bornman, J. F., and Sundby, C. (1999). The chloroplast small heat shock protein undergoes oxidation-dependent conformational changes and may protect plants from oxidative stress. *Cell Stress & Chaperones* **4**, 129-138.
- Hasegawa, P. M., Bressan, R. A., Zhu, J.-K., and Bohnert, H. J. (2000). Plant cellular and molecular responses to high salinity. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **51**, in press.
- Hellmann, H., Funck, D., Rentsch, D., and Frommer, W. B. (2000). Hypersensitivity of an Arabidopsis sugar signaling mutant toward exogenous proline application. *Plant Physiol.* **122**, 357-367.
- Hirt, H. (2000). Connecting oxidative stress, auxin, and cell cycle regulation through a plant mitogen-activated protein kinase pathway. *Proc. Natl. Acad. Sci. U.S.A.* **97**, 2405-2407.
- Holmström, K.-O., Mäntylä, E., Welin, B., Mandal, A., Paiva, E. T., Tunnela, O. E., and Longdesborough, J. (1996). Drought tolerance in tobacco. *Nature* **379**, 683-684.
- Hong, S.-W., and Vierling, E. (2000). Mutants of Arabidopsis thaliana defective in the acquisition of tolerance to high temperature stress. *Proc. Natl. Acad. Sci. U.S.A.* **97**,

4392-4397.

- Ingram, J., and Bartels, D. (1996). The molecular basis of dehydration tolerance in plants. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **47**, 377-403.
- Ishitani, M., Xiong, L., Lee, H., Stevenson, B., and Zhu, J.-K. (1998). HOS1, a genetic locus involved in cold-responsive gene expression in Arabidopsis. *The Plant Cell* **10**, 1151-1161.
- Ishitani, M., Xiong, L., Stevenson, B., and Zhu, J.-K. (1997). Genetic analysis of osmotic and cold stress signal transduction in Arabidopsis: interactions and convergence of abscisic acid-dependent and abscisic acid-independent pathways. *The Plant Cell* **9**, 1935-1949.
- Ismail, A. M., Hall, A. E., and Close, T. J. (1999). Allelic variation of a dehydrin gene cosegregates with chilling tolerance during seedling emergence. *Proc. Natl. Acad. Sci. U.S.A.* **96**, 13566-13570.
- Iturriaga, G., Leyns, L., Villegas, A., Gharaibeh, R., Salamini, F., and Bartels, D. (1996). A family of novel myb-related genes from the resurrection plant *Craterostigma plantagineum* are specifically expressed in callus and roots in response to ABA or desiccation. *Plant Molecular Biology* **32**, 707-716.
- Jaglo-Ottosen, K. R., Gilmour, S. J., Zarka, D. G., Schabenberger, O., and Thomashow, M. F. (1998). Arabidopsis CBF1 overexpression induces COR genes and enhances freezing tolerance. *Science* **280**, 104-106.
- Jain, R. K., and Selvaraj, G. (1997). Molecular genetic improvement of salt tolerance in plants. In "Biotechnology Annual Review" (M. R. El-Gewely, Ed.), Vol. 3, pp. 245-267. Elsevier Science B.V.
- Jonak, C., Kiegl, S., Ligterink, W., Barker, P. J., Huskisson, N. S., and Hirt, H. (1996). Stress signaling in plants: a mitogen-activated protein kinase pathway is activated by cold and drought. *Proceedings of the National Academy of Sciences of the USA* **93**, 11274-11279.
- Kaldenhoff, R., Grote, D., Zhu, J.-J., and Zimmermann, U. (1998). Significance of plasmalemma aquaporins for water-transport in Arabidopsis thaliana. *The Plant Journal* **14**, 121-128.
- Karakas, B., Ozias-Akins, P., Stushnoff, C., Suefferheld, M., and Rieger, M. (1997). Salinity and drought tolerance of mannitol-accumulating transgenic tobacco. *Plant Cell Envir.* **20**, 609-616.
- Kasuga, M., Liu, Q., Miura, S., Yamaguchi-Shinozaki, K., and Shinozaki, K. (1999). Improving plant drought, salt, and freezing tolerance by gene transfer of a single stress-inducible transcription factor. *Nature Biotechnology* **17**, 287-291.
- Kawasaki, S., Borchert, C., Deyholos, M., Wang, H., Brazille, S., Kawai, K., Galbraith, D., and Bohnert, H.J. (2001). Gene expression profiles during the initial phase of salt stress in rice. *The Plant Cell* **13**, 889-905.
- Kaye, C., Neven, L., Hofig, A., Li, Q. B., Haskell, D., and Guy, C. (1998). Characterization of a gene for spinach CAP160 and expression of two spinach cold-acclimation proteins in tobacco. *Plant Physiol.* **116**, 1367-1377.
- Kieber, J. J., Rothenberg, M., Roman, G., Feldmann, K. A., and Ecker, J. R. (1993). CTR1: a negative regulator of the ethylene response pathway in Arabidopsis, encodes a member of the Raf family of protein kinases. *Cell* **72**, 427-441.
- Kingston-Smith, A. H., Harbinson, J. H., and Foyer, C. H. (1999). Acclimation of

- photosynthesis, H₂O₂ content and antioxidants in maize (*Zea mays*) grown at sub-optimal temperatures. *Plant, Cell and Environment* **22**, 1071-1083.
- Kishor, P. B. K., Hong, Z., Miao, G.-H., Hu, C.-A. A., and Verma, D. P. S. (1995). Overexpression of Δ^1 -pyrroline-5-carboxylate synthetase increases proline production and confers osmotolerance in transgenic plants. *Plant Physiol.* **108**, 1387-1394.
- Kiyosue, T., Yamaguchi-Shinozaki, K., and Shinozaki, K. (1994). Cloning of cDNAs for genes that are early-responsive to dehydration stress (ERDs) in *Arabidopsis thaliana* L.: identification of three ERDs as HSP cognate genes. *Plant Mol. Biol.* **25**, 791-798.
- Kjellbom, P., Larsson, C., Johansson, I., Karlsson, M., and Johanson, U. (1999). Aquaporins and water homeostasis in plants. *Trends in Plant Sci.* **4**, 308-314.
- Kodama, H., Hamada, T., Horiguchi, G., Nishimura, M., and Iba, K. (1994). Genetic enhancement of cold tolerance by expression of a gene for chloroplast ω -3 fatty acid desaturase in transgenic tobacco. *Plant Physiol.* **105**, 601-605.
- Kovtun, Y., Chiu, W.-L., Tena, G., and Sheen, J. (2000a). Functional analysis of oxidative stress-activated mitogen-activated protein kinase cascade in plants. *Proc. Natl. Acad. Sci. U.S.A.* **97**, 2940-2945.
- Kovtun, Y., Chiu, W.-L., Tena, G., and Sheen, J. (2000b). Functional analysis of oxidative stress-activated mitogen-activated protein kinase cascade in plants. *Proc. Natl. Acad. Sci. U.S.A.* **97**, 2940-2945.
- Lebreton, C., Jazic-Jancic, V., Steed, A., Pekic, S., and Quarrie, S. A. (1995). Identification of QTL's for drought responses in maize and their use in testing causal relationships between traits. *J. Exp. Botany* **46**, 853-865.
- Lee, G. J., Roseman, A. M., Saibil, H. R., and Vierling, E. (1997). A small heat shock protein stably binds heat-denatured model substrates and can maintain a substrate in a folding-competent state. *EMBO J.* **16**, 659-671.
- Lee, G. J., and Vierling, E. (2000). A small heat shock protein cooperates with heat shock protein 70 systems to reactivate a heat-denatured protein. *Plant Physiol.* **122**, 189-197.
- Lee, J. H., Hubel, A., and Schoffl, F. (1995). Derepression of the activity of genetically engineered heat shock factor causes constitutive synthesis of heat shock proteins and increased thermotolerance in transgenic *Arabidopsis*. *Plant J.* **8**, 603-612.
- Lee, Y. R., Nagao, R. T., and Key, J. L. (1994). A soybean 101-kD heat shock protein complements a yeast HSP 104 deletion mutant in acquiring thermotolerance. *Plant Cell* **6**, 1889-1897.
- Leung, J., and Giraudat, J. (1998). Absciscic acid signal transduction. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **49**, 199-222.
- Lilius, G., Homberg, N., and Bülow, L. (1996). Enhanced NaCl stress tolerance in transgenic tobacco expressing bacterial choline dehydrogenase. *Bio-Technology* **14**, 177-180.
- Lilley, J. M., Ludlow, M. M., McCouch, S. R., and O'Toole, J. C. (1996). Locating QTL for osmotic adjustment and dehydration tolerance in rice. *J. Exp. Botany* **47**, 1427-1436.
- Lippuner, V., Cyert, M. S., and Gasser, C. S. (1996). Two classes of plant cDNA clones differentially complement yeast calcineurin mutants and increase salt tolerance in wild-type yeast. *J. Biol. Chem.* **271**, 12859-12866.
- Liu, J., Ishitani, M., Halfter, U., Kim, C.-S., and Zhu, J.-K. (2000). The *Arabidopsis thaliana*

- SOS2 gene encodes a protein kinase that is required for salt tolerance. *Proc. Natl. Acad. Sci. U.S.A.* **97**, 3730-3734.
- Liu, J., and Zhu, J.-K. (1997a). An *Arabidopsis* mutant that requires increased calcium for potassium nutrition and salt tolerance. *Proc. Natl. Acad. Sci. U.S.A.* **94**, 14960-14964.
- Liu, J., and Zhu, J.-K. (1997b). Proline accumulation and salt-stress-induced gene expression in a salt-hypersensitive mutant of *Arabidopsis*. *Plant Physiol.* **114**, 591-596.
- Liu, J., and Zhu, J.-K. (1998). A calcium sensor homolog required for plant salt tolerance. *Science* **280**, 1943-1945.
- Liu, Q., Kasuga, M., Sakuma, Y., Abe, H., Miura, S., Yamaguchi-Shinozaki, K., and Shinozaki, K. (1998). Two transcription factors, DREB1 and DREB2, with and EREBP/AP2 DNA binding domain separate two cellular signal transduction pathways in drought- and low- temperature-responsive gene expression, respectively, in *Arabidopsis*. *Plant Cell* **10**, 1391-1406.
- Matheos, D. P., Kingsbury, T. J., Ahsan, U. S., and Cunningham, K. W. (1997). Tcn1p/Crz1p, a calcineurin-dependent transcription factor that differentially regulates gene expression in *Saccharomyces cerevisiae*. *Genes and Development* **11**, 3445-3458.
- McConn, S., Hugly, S., Browse, J., and Somerville, C. (1994). *Plant Physiol.* **106**, 1609.
- McKersie, B. D., and Bowley, S. R. (1997). Active oxygen and freezing tolerance in transgenic plants. In "Plant Cold Hardiness. Molecular Biology, Biochemistry and Physiology" (P. H. Li and T. H. H. Chen, Eds.), pp. 203-213. Plenum, New York.
- McKersie, B. D., Bowley, S. R., Harjanto, E., and Leprince, O. (1996). Water-deficit tolerance and field performance of transgenic alfalfa overexpressing superoxide dismutase. *Plant Physiol.* **111**, 1177-1181.
- McKersie, B. D., Bowley, S. R., and Jones, K. S. (1999). Winter survival of transgenic alfalfa overexpressing superoxide dismutase. *Plant Physiol.* **119**, 839-847.
- McKersie, B. D., Chen, Y., Beus, M., Bowley, S. R., Bowler, C., Inze, D., D'Halluin, K., and Botterman, J. (1993). Superoxide Dismutase enhances tolerance of freezing stress in transgenic alfalfa (*Medicago sativa* L.). *Plant Physiology* **103**, 1155-1163.
- McKown, R., Kuroki, G., and Warren, G. (1996). Cold responses of *Arabidopsis* mutants impaired in freezing tolerance. *J. Exp. Botany* **47**, 1919-1925.
- Mizoguchi, T., Ichimura, K., Yoshida, R., and Shinozaki, K. (2000). MAP kinase cascades in *Arabidopsis*: their roles in stress and hormone responses. In "MAP Kinases in Plant Signal Transduction." (H. Hirt, Ed.), Vol. 27, pp. 29-38. Springer-Verlag, Berlin, Heidelberg.
- Møller, S. G., and Chua, N.-H. (1999). Interactions and intersections of plant signaling pathways. *J. Mol. Biol.* **283**, 219-234.
- Murakami, Y., Tsuyama, M., Kobayashi, Y., Kodama, H., and Iba, K. (2000). Trienoic fatty acids and plant tolerance of high temperature. *Science* **287**, 476-479.
- Nakayama, H., Yoshida, K., Ono, H., Murooka, Y., and Shinmyo, A. (2000). Ectoine, the compatible solute of *Halomonas elongata*, confers hyperosmotic tolerance in cultured tobacco cells. *Plant Physiol.* **122**, 1239-1247.
- Nanjo, T., Kobayashi, M., Yoshida, Y., Kakubari, Y., Yamaguchi-Shinozaki, K., and Shinozaki, K. (1999a). Antisense suppression of proline degradation improves tolerance to freezing and salinity in *Arabidopsis thaliana*. *FEBS Letters* **461**, 205-210.

- Nanjo, T., Kobayashi, M., Yoshiba, Y., Sanada, Y., Wada, K., Tsukaya, H., Kakubari, Y., Yamaguchi-Shinozaki, K., and Shinozaki, K. (1999b). Biological functions of proline in morphogenesis and osmotolerance revealed in antisense transgenic *Arabidopsis thaliana*. *Plant J.* **18**, 185-193.
- Niu, X., Bressan, R. A., Hasegawa, P. M., and Prado, J. M. (1995). Ion homeostasis in NaCl stress environments. *Plant Physiol.* **109**, 735-742.
- Noctor, G., and Foyer, C. H. (1998). Ascorbate and glutathione: Keeping active oxygen under control. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **49**, 249-279.
- Nover, L., Scharf, K.-D., Gagliardi, D., Vergne, P., Czarnecka-Verner, E., and Gurley, W. B. (1996). The Hsf world: classification and properties of plant heat stress transcription factors. *Cell Stress & Chaperones* **1**, 215-223.
- Nuccio, M. L., Rhodes, D., McNeil, S. D., and Hanson, A. D. (1999). Metabolic engineering of plants for osmotic stress resistance. *Curr. Opinion in Plant Biol.* **2**, 128-134.
- Nuccio, M. L., Russell, B. L., Nolte, K. D., Rathinasabapathi, B., Gage, D. A., and Hanson, A. D. (1998). The endogenous choline supply limits glycine betaine synthesis in transgenic tobacco expressing choline monooxygenase. *Plant J.* **16**, 487-498.
- Pardo, J. M., Reddy, M. P., Yang, S., Maggio, A., Huh, G.-H., Matsumoto, T., Coca, M. A., Paino-D'Urzo, M., Koiwa, H., Yun, D.-J., Watad, A. A., Bressan, R. A., and Hasegawa, P. M. (1998). Stress signaling through Ca²⁺/calmodulin-dependent protein phosphatase calcineurin mediates salt adaptation in plants. *Proc. Natl. Acad. Sci. U.S.A.* **95**, 9681-9686.
- Parsell, D. A., and Lindquist, S. (1993). The function of heat-shock proteins in stress tolerance: Degradation and reactivation of damaged proteins. *Annu. Rev. Genet.* **27**, 437-496.
- Petrusa, L. M., and Winicov, I. (1997). Proline status in salt-tolerant and salt-sensitive alfalfa cell lines and plants in response to NaCl. *Plant Physiol. Biochem.* **35**, 303-310.
- Pilon-Smits, E. A. H., Ebskamp, M. J. M., Paul, M. J., Jeuken, M. J. W., Weisbeek, P. F., and Smeekens, S. C. M. (1995). Improved performance of transgenic fructan-accumulating tobacco under drought stress. *Plant Physiol.* **107**, 125-130.
- Prändl, R., Hinderhofer, K., Eggers-Schumacher, G., and Schöffl, F. (1998). HSF3, a new heat shock factor from *Arabidopsis thaliana*, derepresses the heat shock response and confers thermotolerance when overexpressed in transgenic plants. *Mol. Gen. Genet.* **258**, 269-278.
- Qin, X., and Zeevaart, J. A. D. (1999). The 9-cis-epoxycarotenoid cleavage reaction is the key regulatory step of abscisic acid biosynthesis in water-stressed bean. *Proc. Natl. Acad. Sci. U.S.A.* **96**, 15354-15361.
- Queitsch, C., Hong, S.-W., Vierling, E., and Lindquist, S. (2000). Heat shock protein 101 plays a crucial role in thermotolerance in *Arabidopsis*. *Plant Cell* **12**, 479-492.
- Ray, I. M., Townsend, M. S., Muncy, C. M., and Henning, J. A. (1999). Heritabilities of water-use efficiency traits and correlations with agronomic traits in water-stressed alfalfa. *Crop Sci.* **39**, 494-498.
- Reichman, J. L., Heard, J., Martin, G., Reuber, L., Jiang, C.-A., Keddie, J., Adam, L., Pineda, O., Tatchell, O. J., Samaha, R. R., Creelman, R., Pilgrim, M., Broun, P., Zhang, J. Z., Grandehari, D., Sherman, B. K., Yu, G.-L. (2000). *Arabidopsis* transcription factors; genome-wide comparative analysis among eukaryotes. *Science*, **290**, 2105-2110.
- Rep, M., Reiser, V., Gartner, U., Thevelein, J. M., Hohmann, S., Ammerer, G., and Ruis, H.

- (1999). Osmotic stress-induced gene expression in *Saccharomyces cerevisiae* requires Msn1p and the novel nuclear factor Hot1p. *Mol. Cell. Biol.* **19**, 5474-5485.
- Ribaut, J.-M., and Hoisington, D. (1998). Marker-assisted selection: new tools and strategies. *Trends in Plant Sci.* **3**, 236-239.
- Roxas, V. P., Smith, R. K., Allen, E. R., and Allen, R. D. (1997). Overexpression of glutathione S-transferase/glutathione peroxidase enhances the growth of transgenic tobacco during stress. *Nature Biotechnology* **15**, 988-991.
- Rubio, F., Gassmann, W., and Schroeder, J. I. (1996). Sodium-driven potassium uptake by the plant potassium transporter HKT1 and mutations conferring salt tolerance. *Science* **270**, 1660-1663.
- Sakamoto, A., Alia, and Murata, N. (1998). Metabolic engineering of rice leading to biosynthesis of glycinebetaine and tolerance to salt and cold. *Plant Mol. Biol.* **38**, 1011-1019.
- Schirmer, E. C., Lindquist, S., and Vierling, E. (1994). An Arabidopsis heat shock protein complements a thermotolerance defect in yeast. *Plant Cell* **6**, 1899-1909.
- Schöfl, F., Prändl, R., and Reindl, A. (1998). Regulation of the heat-shock response. *Plant Physiol.* **117**, 1135-1141.
- Seki, M., Narusaka, M., Abe, H., Kasuga, M., Yamaguchi-Shinozaki, K., Carninici, P., Hayashizaki, Y., and Shinozaki, K. (2001). Monitoring the expression pattern of 1300 Arabidopsis genes under drought and cold stresses by using a full-length cDNA microarray. *The Plant Cell*, **13**, 61-72.
- Serrano, R., Cullianz-Macia, A., and Moreno, V. (1999). Genetic engineering of salt and drought tolerance with yeast regulatory genes. *Scientia Hort.* **78**, 261-269.
- Sheen, J. (1996). Ca^{+2} -Dependent protein kinases and stress signal transduction in plants. *Science* **274**, 1900-1902.
- Shen, B., Jensen, R. G., and Bohnert, H. J. (1997). Increased resistance to oxidative stress in transgenic plants by targeting mannitol biosynthesis to chloroplasts. *Plant Physiol.* **113**, 1177-1183.
- Sheveleva, E., Chmara, W., Bohnert, H. J., and Jensen, R. G. (1997). Increased salt and drought tolerance by D-ononitol production in transgenic *Nicotiana tabacum* L. *Plant Physiol.* **115**, 1211-1219.
- Sheveleva, E. V., Marquez, S., Chmara, W., Zegeer, A., Jensen, R. G., and Bohnert, H. J. (1998). Sorbitol-6-phosphate dehydrogenase expression in transgenic tobacco. *Plant Physiol.* **117**, 831-839.
- Shi, H., Ishitani, M., Kim, C., and Zhu, J.-K. (2000). The *Arabidopsis thaliana* salt tolerance gene SOS1 encodes a putative $\text{Na}^{+}/\text{H}^{+}$ antiporter. *Proc. Natl. Acad. Sci. USA*, **97**, 6896-6901.
- Shinozaki, K., and Yamaguchi-Shinozaki, K. (1997). Gene expression and signal transduction in water-stress response. *Plant Physiology* **115**, 327-334.
- Shinozaki, K., and Yamaguchi-Shinozaki, K. (1999). Molecular responses to drought stress. In "Cold, drought, heat and salt stress in higher plants" (K. Shinozaki and K. Yamaguchi-Shinozaki, Eds.). R.G. Landes Co., Austin, TX.
- Shinozaki, K., Yamaguchi-Shinozaki, K., Liu, Q., Kasuga, M., Ichimura, K., Mizoguchi, T., Urao, T., Miyata, S., Nakashima, K., Shinwari, Z. K., Sakuma, Y., Ito, T., and Seki, M. (1999). Molecular responses to drought stress in plants: regulation of gene expression and signal transduction. In "Plant Responses to Environmental Stress."

- (M. F. Smallwood, C. M. Calvert, and D. J. Bowles, Eds.), pp. 133-143. BIOS Scientific Publishers, Oxford.
- Shinwari, Z. K., Nakashima, K., Miura, S., Kasuga, M., Seki, M., Yamaguchi-Shinozaki, K., and Shinozaki, K. (1998). An *Arabidopsis* gene family encoding DRE/CRT binding proteins involved in low-temperature-responsive gene expression. *Biochem. Biophys. Res. Commun.* **250**, 161-170.
- Singla, S. L., Pareek, A., and Grover, A. (1997). Yeast HSP104 homologue rice HSP110 is developmentally- and stress-regulated. *Plant Sci.* **125**, 211-219.
- Steponkus, P. L. (1984). Role of the plasma membrane in freezing injury and cold acclimation. *Annu. Rev. Plant Physiol.* **35**, 543-584.
- Steponkus, P. L., Uemura, M., and Webb, M. S. (1993). Membrane destabilization during freeze-induced dehydration. *Curr. Topics Plant Physiol.* **10**, 37-47.
- Stockinger, E. J., Gilmour, S. J., and Thomashow, M. F. (1997). *Arabidopsis thaliana* CBF1 encodes an AP2 domain-containing transcriptional activator that binds to the C-repeat/DRE, a cis-acting DNA regulatory element that stimulates transcription in response to low temperature and water deficit. *Proc Natl Acad Sci USA.* **94**, 1035-1040.
- Storlie, E. W., Allan, R. E., and Walker-Simmons, M. K. (1998). Effect of the *Vrn1-Fr1* interval on cold hardiness levels in near-isogenic wheat lines. *Crop Sci.* **38**, 483-488.
- Su, J., Shen, Q., Ho, T.-H. D., and Wu, R. (1998). Dehydration-stress-regulated transgene expression in stably transformed rice plants. *Plant Physiol.* **117**, 913-922.
- Tarczynski, M. C., Jensen, R. G., and Bohnert, H. J. (1993). Stress protection of transgenic tobacco by production of the osmolyte mannitol. *Science* **259**, 508-510.
- Teulat, B., This, D., Khairallah, M., Borries, C., Ragot, C., Sourdille, P., Leroy, P., Monneveux, P., and Charrier, A. (1998). Several QTLs involved in osmotic-adjustment trait variation in barley (*Hordeum vulgare* L.). *Theor. Appl. Gen.* **96**, 688-698.
- Thiele, A., Herold, M., Lenk, I., Quail, P. H., and Gatz, C. (1999). Heterologous expression of *Arabidopsis* phytochrome B in transgenic potato influences photosynthetic performance and tuber development. *Plant Physiol.* **120**, 73-81.
- Thomas, J. C., Sepahi, M., Arendall, B., and Bohnert, H. J. (1995). Enhancement of seed germination in high salinity by engineering mannitol expression in *Arabidopsis thaliana*. *Plant Cell Envir.* **18**, 801-806.
- Thomashow, M. F. (1990). Molecular genetics of cold acclimation in higher plants. *Adv. Genet.* **28**, 99-131.
- Thomashow, M. F. (1998). Role of cold-responsive genes in plant freezing tolerance. *Plant Physiology* **118**, 1-7.
- Thomashow, M. F. (1999). Plant Cold Acclimation: Freezing tolerance genes and regulatory mechanisms. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **50**, 571-599.
- Torsethaugen, G., Pitcher, L. H., Zilinskas, B. A., and Pell, E. J. (1997). Overproduction of ascorbate peroxidase in the tobacco chloroplast does not provide protection against ozone. *Plant Physiol.* **114**, 529-537.
- Trossat, C., Rathinasabapathi, B., Weretilnyk, E. A., Shen, T.-L., Huang, Z.-H., Gage, D. A., and Hanson, A. D. (1998). Salinity promotes accumulation of 3-dimethylsulfoniopropionate and its precursor S-methylmethionine in chloroplasts. *Plant Physiol.* **116**, 165-171.

- Tyystjärvi, E., Riikonen, M., Arisi, A.-C. M., Kettunen, R., Jouanin, L., and Foyer, C. H. (1999). Photoinhibition of photosystem II in tobacco plants overexpressing glutathione reductase and poplars overexpressing superoxide dismutase. *Physiol. Plant.* **105**, 409-416.
- Urao, T., Katagiri, T., Mizoguchi, T., Yamaguchi-Shinozaki, K., Hayashida, N., and Shinozaki, K. (1994). Two genes that encode Ca²⁺-dependent protein kinases are induced by drought and high-salt stresses in *Arabidopsis thaliana*. *Mol. Gen. Genet.* **244**, 331-340.
- Van Camp, W., Capiou, K., Van Montagu, M., Inzé, D., and Slooten, L. (1996). Enhancement of oxidative stress tolerance in transgenic tobacco plants overproducing Fe-superoxide dismutase in chloroplasts. *Plant Physiol.* **112**, 1703-1714.
- van der Krol, A. R., van Poecke, R. M. P., Vorst, O. F. J., Voogt, C., van Leeuwen, W., Borst-Vrengen, T. W. M., Takatsuji, H., and van der Plas, L. H. W. (1999). Developmental and wound-, cold-, desiccation-, ultraviolet-B-stress-induced modulations in the expression of the petunia zinc finger transcription factor gene *ZPT2-2*. *Plant Physiol.* **121**, 1153-1162.
- van der Luit, A. H., Olivari, C., Haley, A., Knight, M. R., and Trewavas, A. J. (1999). Distinct calcium signaling pathways regulate calmodulin gene expression in tobacco. *Plant Physiol.* **121**, 705-714.
- Veena, Reddy, V. S., and Sopory, S. K. (1999). Glyoxalase I from *Brassica juncea*: molecular cloning, regulation and its over-expression confer tolerance in transgenic tobacco under stress. *Plant J.* **17**, 385-395.
- Verma, D. P. S. (1999). Osmotic stress tolerance in plants: role of proline and sulfur metabolisms. In "Molecular Responses to Cold, Drought, Heat and Salt Stress in Higher Plants" (K. Shinozaki and K. Yamaguchi-Shinozaki, Eds.), pp. 153-168. R.G.Landes Co., Austin, TX.
- Vierling, E. (1991). The roles of heat shock proteins in plants. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **42**, 579-620.
- Warren, G., McKown, R., Marin, A., and Teutonico, R. (1996). Isolation of mutations affecting the development of freezing tolerance in *Arabidopsis thaliana* (L.) Heynh. *Plant Physiol.* **111**, 1011-1019.
- Waters, E. R., Lee, G. J., and Vierling, E. (1996). Evolution, structure and function of the small heat-shock proteins in plants. *J. Exp. Bot.* **47**, 325-338.
- Wehmeyer, N., and Vierling, E. (2000). The expression of small heat shock proteins in seeds responds to discrete developmental signals and suggests a general protective role in desiccation tolerance. *Plant Physiol.* **122**, 1099-1108.
- Werner-Fraczek, J. E., and Close, T. J. (1998). Genetic studies of *Triticeae* dehydrins: assignment of seed proteins and a regulatory factor to map positions. *Theor. Appl. Genet.* **97**, 220-226.
- Winicov, I. (1991). Characterization of salt tolerant alfalfa (*Medicago sativa* L.) plants regenerated from salt tolerant cell lines. *Plant Cell Rep* **10**, 561-564.
- Winicov, I. (1993). cDNA encoding putative zinc finger motifs from salt-tolerant alfalfa (*Medicago sativa* L.) cells. *Plant Physiol.* **102**, 681-682.
- Winicov, I. (1996). Characterization of rice (*Oryza sativa* L.) plants regenerated from salt-tolerant cell lines. *Plant Sci.* **113**, 105-111.
- Winicov, I. (1998). New molecular approaches to improving salt tolerance in crop plants.

Annals of Botany **82**, 703-710.

- Winicov, I. (2000a). Alfin1 transcription factor overexpression enhances plant root growth under normal and saline conditions and improves salt tolerance in alfalfa. *Planta* **210**, 416-422.
- Winicov, I. (2000b). Molecular strategies to overcome salt stress in agriculture. In "Molecular tools for the assessment of plant adaptation to the environment" (M. J. Hawkesford and P. Buchner, Eds.), pp. in press. Kluwer, The Netherlands.
- Winicov, I., and Bastola, D. R. (1997). Salt tolerance in crop plants: new approaches through tissue culture and gene regulation. *Acta Physiologia Plant.* **19**, 435-449.
- Winicov, I., and Bastola, D. R. (1999). Transgenic overexpression of the transcription factor Alfin1 enhances expression of the endogenous MsPRP2 gene in alfalfa and improves salinity tolerance of the plants. *Plant Physiol* **120**, 473-80.
- Winicov, I., and Button, J. D. (1991). Induction of photosynthesis gene transcripts by sodium chloride in a salt tolerant alfalfa cell line. *Planta* **183**, 478-483.
- Winicov, I., and Krishnan, M. (1996). Transcriptional and post-transcriptional activation of genes in salt-tolerant alfalfa cells. *Planta* **200**, 397-404.
- Winicov, I., and Shirzadegan, M. (1997). Tissue specific modulation of salt inducible gene expression: callus versus whole plant response in salt tolerant alfalfa. *Physiol Plant* **100**, 314-319.
- Winicov, I., Waterborg, J. H., Harrington, R. E., and McCoy, T. J. (1989). Messenger RNA induction in cellular salt tolerance of alfalfa (*Medicago sativa*). *Plant Cell Rep* **8**, 6-11.
- Xin, Z., and Browse, J. (1998). *eskimo1* mutants of *Arabidopsis* are constitutively freezing-tolerant. *Proc. Natl. Acad. Sci. U.S.A.* **95**, 7799-7804.
- Xu, D., Duan, X., Wang, B., Hong, B., Ho, T.-H. D., and Wu, R. (1996). Expression of a late embryogenesis abundant protein gene, HVA1, from barley confers tolerance to water deficit and salt stress in transgenic rice. *Plant Physiol.* **110**, 249-257.
- Yang, W. J., Nadolska-Orczyk, A., Wood, K. V., Hahn, D. T., Rich, P. J., Wood, A. J., Saneoka, H., Premachandra, G. S., Bonham, C. C., Rhodes, J. C., Joly, R. J., Samaras, Y., Godsbrough, P. B., and Rhodes, D. (1995). Near-isogenic lines of maize differing for glycinebetaine. *Plant Physiology* **107**, 621-630.
- Zhang, H-X., and Blumwald, E. (2001) Transgenic salt-tolerant tomato plants accumulate salt in foliage but not fruit. *Nature Biotechnology*, **19**, 765-768.
- Zhu, B., Su, J., Chang, M., Verma, D. P. S., Fan, Y.-L., and Wu, R. (1998a). Overexpression of a Δ^1 -pyrroline-5-carboxylate synthetase gene and analysis of tolerance to water- and salt-stress in transgenic rice. *Plant Sci.* **139**, 41-48.
- Zhu, J.-K., Liu, J., and Xiong, L. (1998b). Genetic analysis of salt tolerance in *Arabidopsis*: evidence for a critical role of potassium nutrition. *The Plant Cell* **10**, 1181-1191.

8

MOLECULAR MARKERS FOR FLOWERING TIME GENES IN CROP SPECIES

DAVID A. LAURIE AND SIMON GRIFFITHS

Crop Genetics Department

John Innes Centre

Norwich Research Park, Colney, Norwich NR4 7UH, UK

INTRODUCTION

1. The importance of the timing of flowering

Wild plants are highly adapted to the conditions in which they grow. An important component of adaptation is the timing of flowering, which occurs when conditions are most likely to maximize pollination, seed development, seed dispersal and subsequent germination. The importance of reproductive success has resulted in the evolution of sophisticated mechanisms to ensure that the timing of flowering is correct. These mechanisms utilize a range of environmental cues including light quality, day length, temperature and the availability of water and nutrients as well as endogenous signals related to the size and age of the plant. The relative importance of the different components of control varies greatly between species and even between ecotypes.

The timing of flowering is equally important for crop plants. All major crops are grown in regions that extend far beyond the habitats of their wild ancestors and consequently their flowering time characteristics have been adapted, consciously or otherwise, by farmers and plant breeders. For most crops, selection must have been occurring for several thousand years, and consequently all major crops now exist as a range of agritypes (agricultural “ecotypes”) that differ in many characters.

Although plant responses to a range of environmental and internal signals have probably been altered by human selection, there are a number of major effects that are reasonably well characterized. Two of these are plant response to day length (photoperiod response) and response to extended periods of low temperature, typically 4 to 10°C (vernalization response). For photoperiodic response, plant physiologists classify plants as long-day (LD), “which only flower, or flower most rapidly, with more than a certain number of hours of light in each 24h period”; short-day (SD), “which only flower, or which flower most rapidly, with fewer than a certain number of hours of light in each 24h period” or day-neutral (DN), “which flower at the same time irrespective of the photoperiodic conditions” (Thomas and Vince-Prue 1997). An additional level of control is the number of days of appropriate length that are required to induce flowering.

The importance of flowering time to cultivated plants can be illustrated by examples from maize and wheat. Maize is generally believed to have originated in Central America and varieties from this region, and the wild relatives of maize (the teosintes), are quantitative short day plants. A major factor in the expansion of maize cultivation has been the selection of day length insensitive genotypes that can be grown at progressively higher latitudes. Another important factor in species such as maize, where F_1 hybrids form a high proportion of seed sales, is the need to match the flowering times of the parental lines. Photoperiod response is important in the adaptation of sorghum (Craufurd *et al.* 1999) and the selection of day length insensitive rice varieties has been important in the expansion of rice growing areas.

In wheat we can consider two responses. The first concerns sensitivity to photoperiod in European winter wheats (Worland *et al.* 1998; Ferrara *et al.* 1998). Daylength insensitive varieties are earlier flowering and have been shown to have a clear advantage in regions with hot, dry summers because earlier flowering avoids heat stress. Similar correlations were observed in the wild tetraploid wheat *T. dicoccoides* (Kato *et al.* 1998). Daylength insensitivity in cultivated bread wheat can give yield advantages of over 35% in Southern European environments and 15% in Central Europe (Worland 1996). In the UK daylength insensitivity confers yield benefits in hot, dry summers but not in cooler wetter ones. Thus, over a ten year period the yield of daylength insensitive genotypes was found to vary from +9 to -16% (Worland *et al.* 1998), reflecting the year to year vagaries of the British climate.

A second example concerns vernalization response. Spring and winter forms of wheat are widely grown, but each has characteristics making them adapted to specific areas. For example, Iwaki *et al.* (2000) showed that in China, Korea and Japan, spring types are typically grown in areas where winter temperatures are too low (below -7°C) to allow plants to survive from an autumn sowing or in areas of high temperature where vernalization would be an ineffective method of flowering control. Winter types are typically grown where winter temperatures are relatively mild (mean mid winter temperatures of -7 to $+4^{\circ}\text{C}$) and where autumn establishment confers a growth and yield advantage. These parameters probably apply world wide. In the UK, for example, with its relatively mild winters, more than 90% of the wheat acreage is sown with winter types. These examples illustrate the impact of flowering time on cereal crop performance and the importance of understanding how flowering is controlled.

Flowering time may also affect other aspects of crop performance indirectly. For example, photoperiod insensitivity has been shown to affect the severity of yellow rust infection in wheat (Worland and Law 1986) and diseases that infect the developing spike may show different levels of infection in genotypes with different flowering times (Zhu *et al.* 1999).

2. The genetic control of flowering in Arabidopsis

Before we consider flowering time genes in crop species it is useful to summarize briefly what is known about the model dicot *Arabidopsis thaliana*. The experimental advantages of *Arabidopsis* are well documented (Meinke *et al.* 1998; Parinov and Sundaresan 2000) and have enabled researchers to isolate several key genes that regulate flowering. Thus, understanding of the genetic control of flowering time is much more advanced in *Arabidopsis* than in crop species and this provides an excellent basis from which to consider the complexities of the process.

At least 80 *Arabidopsis* genes have been shown to affect flowering and studies of their effects, singly and in combination, have enabled models of the control of flowering to be developed (recently reviewed by Koornneef *et al.* (1998), Levy and Dean (1998), Simpson *et al.* (1999), Reeves and Coupland (2000) and Samach and Coupland (2000)). In these models, flowering is regulated by photoperiod, autonomous, vernalization and gibberellin pathways (Simpson *et al.* 1999).

Day length perception involves red and far red receptors (phytochromes) and blue light receptors (cryptochromes). Light regulates the circadian clock which in turn regulates photoperiodic influences on flowering (Reeves and Coupland 2000; Strayer *et al.* 2000). Suitable day lengths increase expression of genes in the photoperiod pathway including *GIGANTEA* (*GI*) (Fowler *et al.* 1999; Park *et al.* 1999) and *CONSTANS* (*CO*) (Putterill *et al.* 1995). *GI* has characteristics of a membrane protein with intra and extra-cellular domains. *CO* is a transcription factor and experiments involving *CO* under the control of an inducible promoter have identified four target genes including *FLOWERING LOCUS T* (*FT*) (Samach *et al.* 2000).

Recent studies have also identified the regulation of *FLOWERING LOCUS C* (*FLC*) as an important component of flowering time control (Michaels and Amasino 1999; Sheldon *et al.* 1999, 2000). *FLC* represses the transition to flowering and *FLC* expression is promoted by *FRIGIDA* (*FRI*) and reduced by genes of the autonomous pathway of which the *FCA* gene (Macknight *et al.* 1997) has been identified as an important component. *FLC* expression is reduced by vernalization, showing how this signal is integrated with the autonomous pathway. *FLC*, like *CO*, has been implicated in the regulation of *FT* (Samach *et al.* 2000), and this, together with other experiments, is starting to show how different pathways are integrated to control flowering (Samach *et al.* 2000; Devlin and Kay 2000; Blázquez and Weigel, 2000). The flowering time pathways ultimately regulate the expression of the meristem identity genes *LEAFY* (*LFY*) and *APETALA1* (*API*) which in turn regulate genes controlling the structure of flowers (reviewed by Ng and Yanofsky 2000). A considerable amount of work is underway to isolate additional components of the various pathways and understanding of flowering time control in *Arabidopsis* will continue to advance rapidly.

The overall picture from research on *Arabidopsis* is one of a complex series of interlinked pathways that provide a series of checks and balances designed to ensure that the plant flowers at a time that is optimal for reproductive success. Other plants are likely to show variations, perhaps major variations, on these mechanisms and the relative importance of different pathways can be expected to vary between species. For example the gibberellic acid-dependent pathway is considered more important in *Arabidopsis* than in maize (Colasanti and Sundaresan, 2000) or pea (Weller *et al.* 1997a). However, it is likely that many genes will be conserved in other species, that equivalent pathways, and equivalent levels of complexity, will be found in many plants and therefore that

knowledge from *Arabidopsis* will contribute substantially to understanding flowering time control in other species.

3. Two questions about the genetics of flowering time in crop plants

Although the existence of different agritypes in crop plants has long been known, understanding the genetic basis of the differences is still relatively primitive. Genetic analysis can address two questions concerning the control of flowering time. The first is how many genes regulate differences between and within agritypes in any given species? The second asks whether equivalent variation exists in different species. In other words, are differences in flowering time in different species regulated by orthologous genes and does conservation extend to model plants such as *Arabidopsis*? To answer these questions we require detailed genetic maps of the relevant species as well as populations that segregate for differences in flowering time.

4. Methods for mapping flowering time genes

The first genetic maps of all major crops were based on morphological characters. Because few such markers segregated in any given cross, these maps were slow to develop and relatively imprecise. In wheat, however, the availability of aneuploid stocks allowed the chromosome locations of major genes affecting photoperiod or vernalization response to be identified by cytogenetic methods which included the use of monosomic lines, nullisomic/tetrasomic lines and of substitution lines in which a chromosome from a donor variety replaces the homologous chromosome in a recipient variety (Law *et al.* 1987).

Detailed genetic maps began to develop with the establishment of marker systems that utilized variation in DNA sequence. The first genetic maps of crop species were based on Restriction Fragment Length Polymorphism (RFLP), and these have been used to locate several major genes which will be discussed in detail below. More recently, PCR-based markers have become increasingly important, with current mapping efforts focussing on Amplified Fragment Length Polymorphism (AFLP; Vos *et al.* 1995) and Simple Sequence Repeat (SSR) polymorphisms. Future developments are likely to involve markers that assay Single Nucleotide Polymorphism (SNP) as demonstrated by Cho *et al.* (1999). SSR and SNP markers are well suited to automation and high throughput analysis but are less easily used for comparing species. For the latter

application, RFLP probes, particularly cDNAs, have been preferred as these frequently hybridize to DNAs of related species. This approach has established the extent of colinearity between different genomes and allowed the positions of loci controlling specific traits to be compared. The development of comparative maps has recently been reviewed by Gale and Devos (1997) and Schmidt (2000).

Methods that distinguish the methylation state of genes are also needed because several recent studies have suggested that flowering time is related to the methylation status of cells of the shoot apical meristem (SAM). Finnegan *et al.* (2000) have reviewed methylation in relation to development including vernalization response. Another example comes from recent work on the late flowering *fwa* mutation of *Arabidopsis* which was shown to be due to hypomethylation in regions flanking the gene (Koornneef *et al.* 1998; Finnegan *et al.* 2000).

Flowering time can be assessed in the same way as any other phenotypic variable by genetic mapping of segregating populations developed from parents with differing attributes. In cases where genes of large effect are segregating, flowering time in a population may be resolvable into discrete Mendelian classes. In other cases where the trait behaves in a quantitative manner the underlying genes (quantitative trait loci or QTL) must be located by statistical methods. These have recently been reviewed by Jansen (1996) and Kearsey and Farquhar (1998).

The majority of QTL studies to date have used populations of recombinant inbred lines or doubled haploid lines. However, QTL can be difficult to locate in such populations, particularly if several genes are segregating. Therefore, a backcross (BC) approach may be more efficient as this reduces the background genetic variation (Tanksley and Nelson 1996). Backcross derived lines (BDL) may reveal additional genes as well as resolving QTL positions (Rae *et al.* 1999) and BDLs can be refined to the point where the gene of interest is isolated from other variation and behaves as a Mendelian character in crosses with the recurrent parent. This approach has been used to isolate a gene regulating fruit size in tomato (Alpert and Tanksley, 1996; Frary *et al.* 2000) and for flowering time genes in rice which are discussed in more detail in section 5.4. The development of BDLs is greatly facilitated by the use of molecular markers.

5. Mapped flowering time genes in crop species

Because flowering time is so critical to plant performance, it would be expected that varieties of the same agritype would have similar alleles for genes of large effect such as those that determine whether or not the plant has a functional photoperiod or vernalization response. In contrast, variation between genotypes within the same agritype might be expected to be due to genes of relatively small effect, especially if varieties from the same ecogeographical region are compared. Genetic analysis of crop plants supports this hypothesis. Winter barley \times spring barley crosses using Western European germplasm typically segregate for major genes governing photoperiod and vernalization response as well as for a variety of QTL (Laurie *et al.* 1995). Spring \times spring, or winter \times winter barley crosses tend to show only QTL variation (Bezant *et al.* 1996; Thomas *et al.* 1995 and Backes *et al.* 1995, respectively). Many of these QTL do not appear to be directly involved in responses to environmental signals and so have been referred to as earliness “*per se*” (*eps*) or narrow-sense earliness genes.

The relatively subtle effects of flowering time QTL do not mean that they are of no interest. On the contrary, they provide a valuable source of variation for selection within winter and spring gene pools and their relative insensitivity to environmental signals makes them attractive for “fine-tuning” varieties. The importance of flowering time QTL has been emphasized many times (e.g. Worland 1996) but there has been relatively little analysis of these genes to date.

5.1. TRITICEAE SPECIES

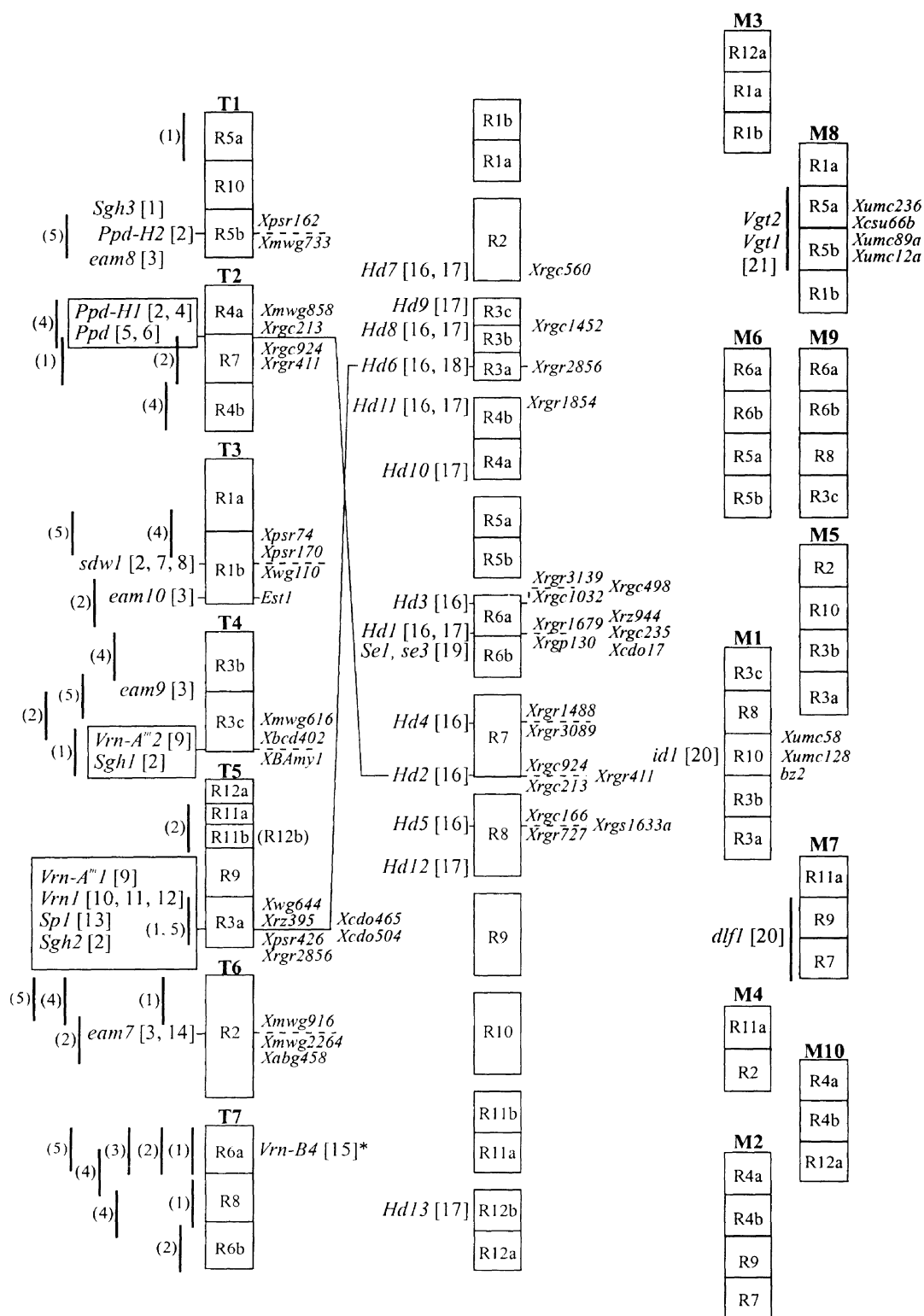
To illustrate the map locations of flowering time genes we compare Triticeae genetic maps (specifically those of wheat, barley and rye) with maps of rice and maize (Fig. 1). This illustrates that although many studies of flowering time genes have been carried out in the Triticeae, analyses of wheat, barley, rye and *T. monococcum* suggest that relatively few genes are involved in distinguishing the photoperiod and vernalization responses of winter and spring types (Fig. 1).

In bread wheat, the only major determinant of photoperiod response that has been identified is the homoeologous *Ppd* gene series mapped on the short arms of the group 2 chromosomes (Börner *et al.* 1998; Worland *et al.* 1998). For these genes, dominant alleles confer insensitivity to day length. Varieties with the day length insensitive allele are earlier flowering and are favoured for wheat

growing areas with hot, dry summers (see section 1). Allelic variation has been identified for all three group 2 chromosomes.

In barley, flowering under long days is promoted by the *Ppd-H1* (*Eam1*) gene on chromosome 2HS (Laurie *et al.* 1995). Comparative mapping using cross-hybridizing RFLP probes shows that the wheat and barley genes are within equivalent chromosome regions, suggesting that day length response is regulated by orthologous genes. A second photoperiod gene in barley (*Ppd-H2*) is located on chromosome 1HL (Fig 1). *Ppd-H2* regulates flowering under short days and has no detectable effect under long days (Laurie *et al.* 1995). *Ppd-H1* and *Ppd-H2* appear to act independently and probably perform very different roles. In winter barleys *Ppd-H2* is proposed to repress flowering under short days and therefore to assist in the adaptation of winter (fall sown) varieties together with vernalization response. *Ppd-H1* then promotes flowering in response to the lengthening days of spring. In spring (spring sown) barleys *Ppd-H2* and vernalization response are unnecessary and a weak response to long days promotes the accumulation of vegetative biomass which in turn provides high yields.

Figure 1. Triticeae (T) and maize (M) chromosomes described in terms of rice linkage blocks. The numbering of the blocks follows Moore *et al.* (1995) but rice chromosomes 1, 3, 4, 11 and 12 have been inverted to give the same orientation as the genetic maps of Harushima *et al.* (1998). Positions of mapped flowering time genes are indicated approximately together with linked RFLP markers (*X* prefix) and in one case an esterase isozyme marker (*Est1*). Numbers in square brackets [] refer to the original articles which should be consulted for details of map locations. [1] Shahla and Tsuchiya (1990); [2] Laurie *et al.* (1995); [3] Franckowiak (1996); [4] Decousset *et al.* (2000); [5] McIntosh *et al.* (1998); [6] Worland *et al.* (1998), Börner *et al.* (1998); [7] Barua *et al.* (1993); [8] Laurie *et al.* (1993); [9] Dubcovsky *et al.* (1998); [10] Galiba *et al.* (1995); [11] Kato *et al.* (1999); [12] Van Deynze *et al.* (1995); [13] Plaschke *et al.* (1993), Börner and Korzun (1998); [14] Stracke and Börner (1998); [15] Chao *et al.* (1989) *N.B. if *Vrn-B4* is translocated from another chromosome it will not be part of the R6a linkage block; [16] Yamamoto *et al.* (1999, 2000); [17] Yano *et al.* (2000); [18] Takahashi *et al.* (1999); [19] Causse *et al.* (1994); [20] Neuffer *et al.* (1997); [21] Vladutu *et al.* (1999). Placement of maize genes with respect to rice linkage blocks also utilized data from Ahn and Tanksley (1993) and Davis *et al.* (1999). Numbers in brackets () refer to QTL from five representative barley studies. (1) Bezant *et al.* (1996); (2) Laurie *et al.* (1995); (3) Backes *et al.* (1995); (4) Hayes *et al.* (1993); (5) Tinker *et al.* (1996). Vertical lines indicate approximate positions and do not represent confidence intervals.



The existence of a mechanism in barley to suppress flowering under short days is supported by the behaviour of the recessive *eam* mutants. These flower early under short days (8 or 10h light), suggesting that the wild type gene product is involved in suppression. The *eam7*, *eam8*, *eam9* and *eam10* mutants (nomenclature following Franckowiak 1996; Franckowiak *et al.* 1996) have been mapped (Fig 1) but to date only *eam7* has been mapped in relation to molecular markers (Stracke and Börner 1998). Similar mutants would be predicted in diploid wheats such as *T. monococcum* but they may not be found in bread wheat unless mutations of all three homoeologous genes can be obtained. Early flowering is also caused by mutation of the *PhyB* gene in barley (Hanumappa *et al.* 1999) and in sorghum (Childs *et al.* 1997) although this variation is not, to our knowledge, used in commercial varieties.

Crosses between winter and spring wheats have identified a homoeologous series of genes that are the principal determinants of vernalization response (McIntosh *et al.* 1998). The genes (*Vrn-A1*, *Vrn-B1* and *Vrn-D1*), which are located on the long arms of the respective group 5 chromosomes, provide different degrees of earliness, and interestingly there is evidence of ecogeographic variation in which alleles are preferred for spring wheat types. For example, Stelmakh (1998) suggests that *Vrn-D1* (*Vrn3*) is more prevalent in equatorial regions. This may be due directly to the characteristics of the genes but it is also possible that these differences may have been influenced by historical patterns of trade and human migration.

A major vernalization response gene has also been mapped on the long arm of the group 5 chromosomes in barley (*Sgh2*), rye (*Sp1*) and the diploid wheat *T. monococcum* (*Vrn-A^m1*). Comparative mapping shows close linkage to the same RFLP markers, suggesting that orthologous genes are involved (Plaschke *et al.* 1993; Galiba *et al.* 1995; Laurie *et al.* 1995; Dubcovsky *et al.* 1998). Two additional vernalization response genes have been described in barley. One is the *Sgh1* gene which has been mapped in relation to RFLP markers on the long arm of chromosome 4H (Laurie *et al.* 1995). Due to chromosome rearrangements that distinguish the 4A and 5A chromosomes from those of B, D and H genomes (Devos *et al.* 1995), this region is colinear with the distal segment of chromosome 5A in wheat where a major gene (*Vrn-A^m2*) has been identified in *T. monococcum* (Dubcovsky *et al.* 1998). This again suggests the involvement of orthologous genes. No equivalent variation is known in bread wheat, probably because early alleles are usually recessive and mutation of the three homoeologous copies may be necessary before the phenotype is seen.

This contrasts with the *Sgh2/Vrn1* loci where dominant alleles confer early flowering and hence are readily detected in bread wheat. However, an additional vernalization gene (*Vrn-B4*) has been mapped on chromosome 7BS of bread wheat (McIntosh *et al.* 1998; Chao *et al.* 1989). This region is involved in a complex rearrangement with chromosomes 4A and 5A (Devos *et al.* 1995) but it is unlikely that *Vrn-B4* corresponds to the *Sgh1/Vrn-A^m2* genes. However, additional mapping data are needed to confirm or disprove this.

A third barley gene (*Sgh3*) is reported to be on the long arm of chromosome 1H (Shahla and Tsuchiya 1990). *Sgh3* resembles *Sgh2/Vrn1* in showing a dominant early phenotype in barley but no equivalent has been positively identified in other species. However, the group 1 chromosomes of bread wheat have been implicated in vernalization response in studies using aneuploid lines (Law *et al.* 1998).

In contrast to the *Ppd* genes, an important element of the vernalization genes is their strong interaction. Thus, winter barleys must have “late” alleles at the *Sgh1*, *Sgh2* and *Sgh3* loci in order to show a strong vernalization response (Takahashi and Yasuda 1971). All other combinations will behave as spring barleys although not all combinations will be equally early flowering. Similar interactions are seen in diploid wheat (Tranquilli and Dubcovsky 2000). Flowering time is also affected by allelic variation at vernalization at these loci.

While major genes affecting photoperiod or vernalization response appear in equivalent locations, QTL differ substantially between crosses and between species. Figure 1 shows the approximate locations of flowering time QTL in five barley studies. These are by no means inclusive of the QTL that have been reported but illustrate the general distribution of QTL over the genome as well as some patterns of clustering, for example on the short arm of 7H. The reason for the latter is unknown. At present QTL comparison between crosses is not meaningful because of the inherent imprecision of QTL location, the possible involvement of a large number of genes and a frequent lack of common markers between crosses. QTL comparison between species is especially problematic.

5.2. MAIZE

In contrast to wheat and barley, maize and rice derive from tropical quantitative short day plants. Maize has a major advantage over Triticeae species in that it contains endogenous transposable elements that can be used for gene isolation

(Maes *et al.* 1999; Walbot 2000). However, much of the most intensively studied maize germplasm is relatively daylength insensitive and perhaps for this reason only three mutations affecting flowering time have been identified; *indeterminate1* (*idl*), *delayed flowering* (*dlf1*) and *leafy* (*Lfy1*) (McSteen *et al.* 2000; Neuffer *et al.* 1997). Of these, *idl* and *dlf1* have been mapped (Fig. 1). The *idl* gene has been cloned by transposon tagging and shown to encode a zinc-finger protein, probably a transcription factor, that is implicated in the transmission of signals regulating flowering from leaves to the shoot apical meristem (Colasanti *et al.* 1998). To date, no equivalent of the *idl* mutant has been identified in other species. Two additional closely linked genes have been mapped on chromosome 8 in crosses utilizing the very early flowering Gaspé Flint (Vladutu *et al.* 1999).

In contrast to the relatively small number of major genes that have been mapped, a large number of QTL affecting flowering time have been detected. These are too numerous to describe here, and as with the Triticeae crosses it is difficult to make meaningful comparisons between crosses and with other species because of uncertainty of the exact location of the genes and differences in the markers used. Nevertheless, this variation provides a valuable source of additional material for flowering time studies.

5.3. RICE

Rice has been the subject of intensive genetic investigation and its small genome size (haploid genome size of approximately 450 megabase pairs (mbp) compared to 120mbp for *Arabidopsis*, 5,400 for barley and 16,000 for bread wheat) has made it the choice for an international effort to discover the sequence of the complete genome (Sasaki and Burr, 2000). In contrast to maize and the Triticeae, therefore, map-based gene isolation in rice is likely to play a major role in gene discovery in the near future. Transposon tagging using exogenous (Izawa *et al.* 1997) or endogenous (Hirochika 1997) elements is also under development and rice is also readily transformed by biolistic or *Agrobacterium* methods (Christou 1997; Hiei *et al.* 1997), making it feasible to assay the functions of candidate genes.

At least 23 major genes and numerous QTL affecting flowering time have been reported in rice (Yamamoto *et al.* 2000), although not all have been mapped in relation to molecular markers. The mapped genes are variously named including *Se* and *Hd*. The analysis of the *Hd* genes shows how flowering time effects, which initially may be detected as QTL, can be resolved into Mendelian

factors by backcrossing to appropriate genetic backgrounds (Yano *et al.* 1997; Yamamoto *et al.* 1999, 2000). Yamamoto *et al.* (2000) have demonstrated epistasis between *Hd6* and *Hd2*, showing that, as in LD cereals, gene interactions are important in the regulation of flowering time in rice.

In contrast to the situation in the Triticeae, several flowering time genes have already been cloned from rice. A major determinant of day length sensitivity is the *Hd1* gene on rice chromosome 6 (Yamamoto *et al.* 1999) which is probably the *Se1* gene previously described in the same region. *Hd1* has recently been isolated by a map-based approach (Yano *et al.* 2000) and shown to be homologous to the *Arabidopsis* *CONSTANS* gene. This is particularly interesting because it provides the first molecular evidence that similar genes are used to regulate flowering in LD and SD plants. It suggests that genes such as *CO* may be upregulated in response to inductive day lengths and that differences between LD and SD plants are more to do with what determines an inductive day length. *Hd6* has also been isolated by a map-based approach (Takahashi *et al.* 2001) and shown to encode the alpha subunit of casein kinase II. This gene is also known to affect flowering in *Arabidopsis*, although its role in phosphorylation suggests it may have additional functions (Sugano *et al.* 1999).

An alternative approach has recently been used to isolate the *Se5* gene of rice (Izawa *et al.* 2000). In this case the *se5* mutant was extremely early flowering and biochemical evidence suggested that this was due to loss of functional phytochromes. This in turn suggested that the mutation might be in the rice homologue of the *Arabidopsis* *HY1* gene. Sequencing and transgenic analysis confirmed that the rice *se5* mutant carried a lesion in the heme oxygenase enzyme necessary for the synthesis of the chromophore component of phytochromes. This also established phytochromes as an important component of flowering time in rice. However, comparison of *se5* and *hyl* mutants suggested that phytochromes do not act in exactly the same way in these two species.

5.4. BRASSICA SPECIES

The *Brassica* and *Arabidopsis* genera are part of the Cruciferae family and knowledge from *Arabidopsis* is therefore readily transferred to this group of crop plants. Genetic maps have been aligned but comparisons are made more complex by the polyploidy of several important species such as *B. napus* and by internal duplications within the *Brassica* genomes (Lagercrantz 1998; Lan *et al.*

2000). Nevertheless, colinear regions containing flowering time genes have been identified (Lagercrantz *et al.* 1996; Osborn *et al.* 1997; Bohuon *et al.* 1998). In Osborn *et al.* (1997), for example, chromosome regions affecting vernalization responsive flowering were hypothesised to correspond to the *FLC* and *FRI* genes of *Arabidopsis* (see section 2). Furthermore, the evolutionary similarity facilitates gene isolation from brassicas in cases where the sequence of the *Arabidopsis* gene is known. This is illustrated by studies of a *Brassica napus* gene highly homologous to *CO* which was shown to complement the *co* mutation of *Arabidopsis*, proving the functional equivalence of the proteins (Robert *et al.* 1999).

5.5. PEA

Studies of the control of flowering in pea are assisted by a range of mutants and the suitability of this species for grafting experiments. This has led to the development of a model in which the promotion of flowering by long days is due to a reduction of a graft-transmissible inhibitor and the production of a flowering stimulus (Weller *et al.* 1997a,b). Mutations of the *PhyA* and *PhyB* genes have important effects. Mutants of *phyA* are late flowering and are thought to be unable to down regulate inhibitor synthesis while *phyB* mutants are early flowering, probably because of a failure to activate inhibitor synthesis (Reid *et al.* 1996; Weller *et al.* 1997a). The *Sn*, *Ppd* and *Dne* genes are thought to act in the inhibitor synthesis pathway while the *Gigas (gi)* mutant is likely to be defective in a floral stimulus (Beveridge and Murfet 1996; Reid *et al.* 1996). Grafting experiments were used to test the order of gene action within the inhibitor synthesis pathways. These experiments suggested *Sn* - *Ppd* - *Dne* as the most probable order (Taylor and Murfet 1996).

6. Colinearity and common pathways

As shown in Figure 1, there are intriguing similarities in the genetic map locations of major photoperiod and vernalization genes in the Triticeae. The simplest explanation is that these species share common mechanisms of response and that differences in phenotype are due to allelic variation in orthologous genes, although there is no molecular proof of this. Given the large number of genes involved in flowering in model systems this is a surprising finding. It implies that few genes control photoperiod and vernalization response, that genes controlling these responses are clustered, or that few genes

in the relevant pathways can be altered without making the plants unusable in cultivation. Of the three, the last possibility seems the most plausible.

Figure 1 also shows two examples where flowering time genes apparently coincide in LD and SD cereals. Common pathways are certainly possible, as described above in relation to the identification of the rice *Hd1* gene as a *CONSTANS* homologue. Thus, the colinearity of the map regions containing *Ppd* (Triticeae 2S) and *Hd2* (Rice 7L) or of *Vrn1* and *Hd6* are of interest for detailed comparison. However, models for gene similarity based on comparative mapping need to be used with great caution. There are potentially a large number of genes that can affect flowering and equivalent map locations might therefore arise by chance. In these examples the relationship between *Ppd* and *Hd2* remains unproven for the present, but *Vrn1* and *Hd6* are known to be unrelated. As described above (section 5.3) the *Hd6* gene encodes the casein kinase II α subunit and a mapped clone corresponding to this gene (R2586) has previously been shown to be distal to the wheat *Vrn1* locus, although closely linked (Kato *et al.* 1999). Thus, although the mapping data shows that these two regions are colinear (Sarma *et al.* 1998; Kato *et al.* 1999), high resolution mapping of the relevant genes is essential before meaningful comparisons can be made.

7. Implications for identifying new markers for flowering time genes

Mapping of flowering time genes in relation to RFLP markers is clearly a workable method and a consequence of colinearity is that RFLP markers mapped in other crosses or species become available for use. However, mapping methods are moving rapidly towards PCR or chip-based assays that are designed to be genome, and often allele, specific. This is illustrated by recent work in barley where sequence-tagged-site (STS) markers flanking the *Ppd-H1* gene were developed (Decousset *et al.* 2000). The approach was to develop a BC₂ population in which *Ppd-H1* (*Eam1*) was the only major gene segregating. Early and late flowering plants were then used to make pooled DNA samples that were screened to identify AFLP markers that only appeared in the early pool. As other regions of the genome will be represented in both pools, this identified linked markers. Six AFLP markers were found that were more closely linked than previously identified RFLP markers and the most closely linked were less than 1 cM from the *Ppd-H1* gene. Three of the polymorphic AFLP bands were cloned, sequenced and converted to STS markers. RFLP analysis showed that all three of the cloned bands were low copy sequences in

the donor barley genome. Interestingly, two were absent from the recurrent parent and none hybridized well to DNA of other cereals. Thus, the PCR markers were specific to parts of the barley gene pool and were not transferable to other species.

High throughput systems in the major crop species are likely to focus on markers that are genome, or allele, specific. This may appear disadvantageous for comparative analysis. However, by developing assays for homologous genes in different species it will still be possible to obtain comparative mapping information. This means that it will be possible to advance mapping in each species rapidly and to integrate information which will assist the identification of common regulators of flowering time control.

8. Methods for isolating additional genes controlling flowering in crop plants

Linked markers can be used for selection of specific flowering time alleles, as in the barley example in section 7, but because their efficiency depends on the tightness of linkage they will be most efficient for genes that are accurately mapped. QTL will be more difficult to select with any accuracy unless their map locations are resolved by more detailed study. But linked markers will never be as efficient as selecting the gene itself. Furthermore, without isolating the gene we will have limited understanding of how flowering time control operates and therefore whether particular alleles are suitable for use in individual plant breeding programmes. There will therefore be continuing interest in isolating flowering time genes. As well as furthering understanding of the functional significance of allelic variation, specific alleles could be selected with high efficiency and germplasm collections could be screened for novel variants. Isolation of the genes would also allow transgenic approaches for the regulation of flowering time, although it remains to be seen whether such varieties would gain consumer acceptance. As described in previous sections, several flowering time genes have already been isolated from crop species. However, as the majority remain unknown it is useful to consider strategies for their isolation.

Worldwide there has been large scale investment in the public and private sectors to provide genomic resources for major crop species. These resources are most advanced in rice and make map-based gene isolation easiest in this species. However, this approach is already becoming feasible even in large genome species such as barley, as illustrated by the recent isolation of the *mlo*

and *rar1* genes which are involved in the resistance of barley to powdery mildew (Büschges *et al.* 1997; Shirasu *et al.* 1999).

The maize genome is about five times the size of rice and half that of barley and therefore chromosome walking strategies are feasible. Maize, however, has the advantage of endogenous transposable element systems that can be used to isolate genes based on mutant phenotype or to identify mutants corresponding to known gene sequences. Transposons can also generate novel functional alleles on excision which reveal additional information on gene function. Transposon tagging will feature extensively in future work in maize and is likely to be developed extensively in other species.

There are well established common principles that link crop and model species and in several cases orthologous genes have been shown to regulate equivalent developmental processes. Examples include many “housekeeping” genes, the use of phytochromes for light perception (section 5), and conservation of genes involved in gibberellic acid control of plant growth (Peng *et al.* 1997, 1999). Therefore it is likely that many genes in crop species can be isolated by virtue of their sequence homology to genes from model species such as *Arabidopsis*. We are especially interested in this area and have recently identified barley genes that are related to the *Arabidopsis* *CONSTANS*, *GIGANTEA* and *TERMINAL FLOWER1* genes. Although genetic mapping shows that these genes do not correspond to known flowering time genes in barley, this does suggest that similar pathways of control exist. The current lack of correspondence between the genes and the flowering time loci may simply reflect a lack of appropriate allelic variation in crop species. An additional factor may be differences in the ways in which variation is recognized. In *Arabidopsis*, much effort has been devoted to mutation screens and plants with the clearest phenotypes will tend to be selected. In crop plants, however, most variation has been identified by intercrossing established agritypes or by mutagenesis designed to identify variants with agronomically useful features. Therefore, the spectrum of mutations may differ between crop and model systems. This situation is likely to change with further mutagenesis of crop species and analyses of flowering time variation in natural ecotypes of *Arabidopsis* (Alonso-Blanco *et al.* 1998; Swarup *et al.* 1999; Alonso-Blanco and Koornneef 2000).

Future work will lead to a much greater understanding of the control of flowering in model and crop species, including an elucidation of the differences between LD, SD and DN plants. It will also lead to the provision of many more

markers for the selection of flowering time and the identification of new sources of variation for plant breeding.

Acknowledgements

This work was supported by grant in aid to the John Innes Centre from the Biotechnology and Biological Sciences Research Council of Great Britain. SG acknowledges the support of a John Innes Foundation Studentship. We thank many of our colleagues for valuable discussion and information. Any remaining errors or omissions are our own.

References

- Ahn, S. and Tanksley, S.D. (1993) Comparative linkage maps of the rice and maize genomes, *Proc. Natl. Acad. Sci. USA* **90**, 7980-7984.
- Alonso-Blanco, C., El-Assal, S.E., Coupland, G. and Koornneef, M. (1998) Analysis of natural allelic variation at flowering time loci in the Landsberg *erecta* and Cape Verde Islands ecotypes of *Arabidopsis thaliana*, *Genetics* **149**, 749-764.
- Alonso-Blanco, C. and Koornneef, M. (2000) Naturally occurring variation in *Arabidopsis*: an underexploited resource for plant genetics, *Trends Plant Sci.* **5**, 22-29.
- Alpert, K.B. and Tanksley, S.D. (1996) High-resolution mapping and isolation of a yeast artificial chromosome contig containing fw2.2: A major fruit weight quantitative trait locus in tomato, *Proc. Natl. Acad. Sci. USA* **93**, 15503-15507.
- Backes, G., Graner, A., Foroughi-Weir, B., Fischbeck, G., Wenzel, G. and Jahoor, A. (1995) Localization of quantitative trait loci (QTL) for agronomic important characters by the use of a RFLP map in barley (*Hordeum vulgare* L.), *Theor. Appl. Genet.* **90**, 294-302.
- Barua, U.M., Chalmers, K.J., Thomas, W.T.B., Hackett, C.A., Lea, V., Jack, P., Forster, B.P., Waugh, R. and Powell, W. (1993) Molecular mapping of genes determining height, time to heading, and growth habit in barley (*Hordeum vulgare*), *Genome* **36**, 1080-1087.
- Beveridge, C.A. and Murfet, I.C. (1996) The *gigas* mutant in pea is deficient in the floral stimulus, *Physiol. Plant.* **96**, 637-645.
- Bezant, J.H., Laurie, D.A., Pratchett, N., Chojecki, J. and Kearsey, M.J. (1996) Marker regression mapping of QTL controlling flowering time and plant height in a spring barley (*Hordeum vulgare* L.) cross, *Heredity* **77**, 64-73.
- Blázquez, M.A. and Weigel, D. (2000) Integration of floral inductive signals in *Arabidopsis*, *Nature* **404**, 889-892.
- Bohuon, E.J.R., Ramsay, L.D., Craft, J.A., Arthur, E.A., Marshall, D.F., Lydiate, D.J. and Kearsey, M.J. (1998) The association of flowering time quantitative trait loci with duplicated regions and candidate loci in *Brassica oleracea*, *Genetics* **150**, 393-401.
- Börner, A. and Korzun, V. (1998) A consensus linkage map of rye (*Secale cereale* L.) including 374 RFLPs, 24 isozymes and 15 gene loci, *Theor. Appl. Genet.* **97**, 1279-1288.
- Börner, A., Korzun, V. and Worland, A.J. (1998) Comparative genetic mapping of loci affecting plant height and development in cereals, *Euphytica* **100**, 245-248 1998.

- Büschges, R., Hollricher, K., Panstruga, R., Simons, G., Wolter, M., Frijters, A., van Daelen, R., van der Lee, T., Diergaarde, P., Groenendijk, J., Töpsch, S., Vos, P., Salamini, F. and Schulze-Lefert, P. (1997) The barley *mlo* gene: a novel control element of plant pathogen resistance, *Cell* **88**, 695-705.
- Causse, M.A., Fulton, T.M., Cho, Y.G., Ahn, S.N., Chunwongse, J., Wu, K., Xiao, J., Yu, Z., Ronald, P.C., Harrington, S.E., Second, G., McCouch, S.R. and Tanksley, S.D. (1994) Saturated molecular map of the rice genome based on an interspecific backcross population, *Genetics* **138**, 1251-1274.
- Chao, S., Sharp, P.J., Worland, A.J., Koebner, R.M.D. and Gale, M.D. (1989) RFLP-based genetic maps of wheat homoeologous group 7 chromosomes, *Theor. Appl. Genet.* **78**, 495-504.
- Childs, K.L., Miller, F.R., Cordonnier Pratt, M.M., Pratt, L.H., Morgan, P.W. and Mullet, J.E. (1997) The sorghum photoperiod sensitivity gene, *Ma(3)*, encodes a phytochrome B, *Plant Physiol.* **113**, 611-619.
- Cho, R.J., Mindrinos, M., Richards, D.R., Sapolsky, R.J., Anderson, M., Drenkard, E., Dewdney, L., Reuber, T.L., Stammers, M., Federspiel, N., Theologis, A., Yang, W.H., Hubbell, E., Au, M., Chung, E.Y., Lashkari, D., Lemieux, B., Dean, C., Lipshutz, R.J., Ausubel, F.M., Davis, R.W. and Oefner, P.J. (1999) Genome-wide mapping with biallelic markers in *Arabidopsis thaliana*, *Nature Genetics* **23**, 203-207.
- Christou, P. (1997) Rice transformation: bombardment. *Plant Mol. Biol.* **35**, 197-203.
- Colasanti, J., Yuan, Z. and Sundaresan, V. (1998) The indeterminate gene encodes a zinc finger protein and regulates a leaf-generated signal required for the transition to flowering in maize, *Cell* **93**, 593-603.
- Colasanti, J. and Sundaresan, V. (2000) 'Florigen' enters the molecular age: long-distance signals that cause plants to flower, *Trends in Biochem. Sci.* **25**, 236-240.
- Craufurd, P.Q., Mahalakshmi, V., Bidinger, F.R., Mukuru, S.Z., Chantreau, J., Omanga, P.A., Qi, A., Roberts, E.H., Ellis, R.H., Summerfield, R.J. and Hammer, G.L. (1999) Adaptation of sorghum: characterisation of genotypic flowering responses to temperature and photoperiod, *Theor Appl Genet* **99**, 900-911.
- Davis, G.L., McMullen, M.D., Baysdorfer, C., Musket, T., Grant, D., Staebell, M., Xu, G., Polacco, M., Koster, L., Melia-Hancock, S., Houchins, K., Chao, S. and Coe, E.H. (1999) A maize map standard with sequenced core markers, grass genome reference points and 932 expressed sequence tagged sites (ESTs) in a 1736-locus map, *Genetics* **15**, 1137-1172.
- Decousset, L., Griffiths, S., Dunford, R.P., Pratchett, N. and Laurie, D.A. (2000) Development of STS markers closely linked to the *Ppd-H1* photoperiod response gene of barley (*Hordeum vulgare* L.), *Theor. Appl. Genet.* (in press).
- Devlin, P.F. and Kay, S.A. (2000) Flower arranging in *Arabidopsis*, *Science* **288**, 1600-1602.
- Devos, K.M., Dubcovsky, J., Dvorak, J., Chinoy, C.N. and Gale, M.D. (1995) Structural evolution of wheat chromosomes 4A, 5A, and 7B and its impact on recombination, *Theor. Appl. Genet.* **91**, 282-288.
- Dubcovsky, J., Lijavetzky, D., Appendino, L. and Tranquilli, G. (1998) Comparative RFLP mapping of *Triticum monococcum* genes controlling vernalization requirement, *Theor. Appl. Genet.* **97**, 968-975.

- Ferrara, G.O., Mosaad, M.G., Mahalakshmi, V. and Rajaram, S. (1998) Photoperiod and vernalisation response of Mediterranean wheats, and implications for adaption, *Euphytica* **100**, 377-384.
- Finnegan, E.J., Peacock, W.J. and Dennis, E.S. (2000) DNA methylation, a key regulator of plant development and other processes, *Curr. Opin. Genet. Dev.* **10**, 217-223.
- Fowler, S., Lee, K., Onouchi, H., Samach, A., Richardson, K., Morris, B., Coupland, G. and Putterill, J. (1999) *GIGANTEA*: a circadian clock-controlled gene that regulates photoperiodic flowering in *Arabidopsis* and encodes a protein with several possible membrane-spanning domains, *EMBO J.* **18**, 4679-4688.
- Franckowiak, J.D., Lundqvist, U. and Konishi, T. (1996) New and revised names for barley genes, *Barley Genet. Newsl.* **26**, 4-8.
- Franckowiak, J.D. (1996) Revised linkage maps for morphological markers in barley, *Hordeum vulgare*, *Barley Genet. Newsl.* **26**, 9-21.
- Frary, A., Nesbitt, T.C., Frary, A., Grandillo, S., van der Knaap, E., Cong, B., Liu, J.P., Meller, J., Elber, R., Alpert, K.B. and Tanksley, S.D. (2000) fw2.2: A quantitative trait locus key to the evolution of tomato fruit size, *Science* **289**, 85-88.
- Gale, M.D. and Devos, K.M. (1997) Plant comparative mapping after 10 years, *Science* **282**, 656-659.
- Galiba, G., Quarrie, S.A., Sutka, J., Morgounov, A. and Snape, J.W. (1995) RFLP mapping of the vernalization (*Vrn1*) and frost resistance (*Fr1*) genes on chromosome 5A of wheat, *Theor. Appl. Genet.* **90**, 1174-1179.
- Hanumappa, M., Pratt, L.H., Cordonnier-Pratt, M.M. and Deitzer, G.F. (1999) A photoperiod-insensitive barley line contains a light-labile phytochrome B, *Plant Physiol.* **119**, 1033-1039.
- Harushima, Y., Yano, M., Shomura, A., Sato, M., Shimano, T., Kuboki, Y., Yamamoto, T., Lin, S.Y., Antonio, B.A., Parco, A., Kajiya, H., Huang, N., Yamamoto, K., Nagamura, Y., Kurata, N., Khush, G.S. and Sasaki, T. (1998) A high-density rice genetic map with 2275 markers using a single F₂ population, *Genetics* **148**, 479-494.
- Hayes, P.M., Liu, B.H., Knapp, S.J., Chen, F., Jones, B., Blake, T., Franckowiak, J., Rasmusson, D., Sorrells, M., Ullrich, S.E., Wesenberg, D. and Kleinhofs, A. (1993) Quantitative trait locus effects and environmental interaction in a sample of North American barley germ plasm, *Theor. Appl. Genet.* **87**, 392-401.
- Hiei, Y., Komari, T. and Kubo, T. (1997) Transformation of rice mediated by *Agrobacterium tumefaciens*, *Plant Mol. Biol.* **35**, 205-218.
- Hirochika, H. (1997) Retrotransposons of rice: their regulation and use for genome analysis, *Plant Mol. Biol.* **35**, 231-240.
- Izawa, T., Ohnishi, T., Nakano, T., Ishida, N., Enoki, H., Hashimoto, H., Itoh, K., Terada, R., Wu, C.Y., Miyazaki, C., Endo, T., Iida, S. and Shimamoto, K. (1997) Transposon tagging in rice, *Plant Mol. Biol.* **35**, 219-229.
- Izawa, T., Oikawa, T., Tokutomi, S., Okuno, K. and Shimamoto, K. (2000) Phytochromes confer the photoperiodic control of flowering in rice (a short day plant), *Plant J.* **22**, 391-399.
- Iwaki, K., Nakagawa, K., Kuno, H. and Kato, K. (2000) Ecogeographical differentiation in east Asian wheat, revealed from the geographical variation of growth habit and *Vrn* genotype, *Euphytica* **111**, 137-143.

- Jansen, R.C. (1996) Complex plant traits: Time for polygenic analysis. *Trends Plant Sci.* **1**, 89-94.
- Kato, K., Tanizoe, C., Beiles, A. and Nevo, E. (1998) Geographical variation in heading traits in wild emmer wheat, *Triticum dicoccoides*. II. Variation in heading date and adaptation to diverse eco-geographical conditions, *Hereditas* **128**, 33-39.
- Kato, K., Miura, H. and Sawada, S. (1999) Comparative mapping of the wheat *Vrn-A1* region with the rice *Hd-6* region, *Genome* **42**, 204-209.
- Kearsey, M.J. and Farquhar, A.G.L. (1998) QTL analysis in plants; where are we now? *Heredity* **80**, 137-142.
- Koornneef, M., Alonso-Blanco, C., Peeters, A.J.M. and Soppe, W. (1998) Genetic control of flowering time in *Arabidopsis*. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **49**, 345-370.
- Lagercrantz, U., Putterill, J., Coupland, G. and Lydiate, D. (1996) Comparative mapping in *Arabidopsis* and *Brassica*, fine scale genome collinearity and congruence of genes controlling flowering time, *Plant J.* **9**, 13-20.
- Lagercrantz, U. (1998) Comparative mapping between *Arabidopsis thaliana* and *Brassica nigra* indicates that *Brassica* genomes have evolved through extensive genome replication accompanied by chromosome fusions and frequent rearrangements, *Genetics* **150**, 1217-1228.
- Lan, T.H., DelMonte, T.A., Reischmann, K.P., Hyman, J., Kowalski, S.P., McFerson, J., Kresovich, S. and Paterson, A.H. (2000) An EST-enriched comparative map of *Brassica oleracea* and *Arabidopsis thaliana*, *Genome Res.* **10**, 776-788.
- Laurie, D.A., Pratchett, N., Romero, C., Simpson, E. and Snape, J.W. (1993). Assignment of the *denso* dwarfing gene to the long arm of chromosome 3(3H) of barley by use of RFLP markers, *Plant Breed.* **111**, 198-203.
- Laurie, D.A., Pratchett, N., Bezant, J.H. and Snape, J.W. (1995) RFLP mapping of five major genes and eight quantitative trait loci controlling flowering time in a winter x spring barley (*Hordeum vulgare* L.) cross, *Genome* **38**, 575-585.
- Law, C.N., Snape, J.W. and Worland, A.J. (1987) Aneuploidy in wheat and its uses in genetic analysis, in F.G.H. Lupton (ed), *Wheat Breeding, its Scientific Basis*, Chapman and Hall, London, pp 71-108.
- Law, C.N., Suarez, E., Miller, T.E. and Worland, A.J. (1998) The influence of the group 1 chromosomes of wheat on ear-emergence times and their involvement with vernalization and day length, *Heredity* **80**, 83-91.
- Levy, Y.Y. and Dean, C. (1998) Control of flowering time, *Curr. Opin. Plant Biol.* **1**, 49-54.
- Macknight, R., Bancroft, I., Page, T., Lister, C., Schmidt, R., Love, K., Westphal, L., Murphy, G., Sherson, S., Cobbett, C. and Dean, C. (1997) *FCA*, a gene controlling flowering time in *Arabidopsis*, encodes a protein containing RNA-binding domains, *Cell* **89**, 737-745.
- Maes, T., De Keuleleire, P. and Gerats, T. (1999) Plant tagnology, *Trends Plant Sci.* **4**, 90-96.
- McIntosh, R.A., Hart, G.E., Devos, K.M., Gale, M.D. and Rogers, W.J. (1998) Catalogue of wheat gene symbols. Volume 5, Proc. 9th Int. Wheat Genet. Symp., University Extension Press, University of Saskatchewan.
- McSteen, P., Laudencia-Chingcuanco, D. and Colasanti, J. (2000) A floret by any other name: control of meristem identity in maize, *Trends Plant Sci.* **5**, 61-66.
- Meinke, D.W., Cherry, J.M., Dean, C., Rounsley, S.D. and Koornneef, M. (1998) *Arabidopsis thaliana*: A model plant for genome analysis, *Science* **282**, 662-682.

- Michaels, S.D. and Amasino, R.M. (1999) *FLOWERING LOCUS C* encodes a novel MADS domain protein that acts as a repressor of flowering, *Plant Cell* **11**, 949-956.
- Moore, G., Devos, K.M., Wang, Z. and Gale, M.D. (1995). Grasses, line up and form a circle. *Curr. Biol.* **5**, 737-739.
- Neuffer, M.G., Coe, E.H. and Wessler, S.R. (1997) *Mutants of maize*, Cold Spring Harbor Laboratory Press.
- Ng, M. and Yanofsky, M.F. (2000) Three ways to learn the ABCs, *Curr. Opin. Plant Biol.* **3**, 47-52.
- Osborn, T.C., Kole, C., Parkin, I.A.P., Sharpe, A.G., Kuiper, M., Lydiate, D.J. and Trick, M. (1997) Comparison of flowering time genes in *Brassica rapa*, *B. napus* and *Arabidopsis thaliana*, *Genetics* **146**, 1123-1129.
- Parinov, S. and Sundaresan, V. (2000) Functional genomics in *Arabidopsis*: large-scale insertional mutagenesis complements the genome sequencing project, *Curr. Opin. Biotechnol.* **11**, 157-161.
- Park, D.H., Somers, D.E., Kim, Y.S., Choy, Y.H., Lim, H.K., Soh, M.S., Kim, H.J., Kay, S.A. and Nam, H.G. (1999) Control of circadian rhythms and photoperiodic flowering by the *Arabidopsis* *GIGANTEA* gene, *Science* **285**, 1579-1582.
- Peng, J., Carol, P., Richards, D.E., King, K.E., Cowling, R.J., Murphy, G.P. and Harberd, N.P. (1997) The *Arabidopsis* *GAI* gene defines a signalling pathway that negatively regulates gibberellin responses, *Genes & Dev.* **11**, 3194-3205.
- Peng, J.R., Richards, D.E., Hartley, N.M., Murphy, G.P., Devos, K.M., Flintham, J.E., Beales, J., Fish, L.J., Worland, A.J., Pelica, F., Sudhakar, D., Christou, P., Snape, J.W., Gale, M.D. and Harberd, N.P. (1999) 'Green revolution' genes encode mutant gibberellin response modulators, *Nature* **400**, 256-261.
- Plaschke, J., Börner, A., Xie, D.X., Koebner, R.M.D., Schlegel, R. and Gale, M.D. (1993) RFLP mapping of genes affecting plant height and growth habit in rye, *Theor. Appl. Genet.* **85**, 1049-1054.
- Putterill, J., Robson, F., Lee, K., Simon, R. and Coupland, G. (1995) The *CONSTANS* gene of *Arabidopsis* promotes flowering and encodes a protein showing similarities to zinc finger transcription factors, *Cell* **80**, 847-858.
- Rae, A.M., Howell, E.C. and Kearsley, M.J. (1999) More QTL for flowering time revealed by substitution lines in *Brassica oleracea*, *Heredity* **83**, 586-596.
- Reeves, P.H. and Coupland, G. (2000) Response of plant development to environment: control of flowering by daylength and temperature, *Curr. Opin. Plant Biol.* **3**, 37-42.
- Reid, J.B., Murfet, I.C., Singer, S.R., Weller, J.L. and Taylor, S.A. (1996) Physiological-genetics of flowering in *Pisum*, *Semin. Cell Dev. Biol.* **7**, 445-463.
- Robert, L.S., Robson, F., Sharpe, A., Lydiate, D. and Coupland, G. (1999) Conserved structure and function of the *Arabidopsis* flowering time gene *CONSTANS* in *Brassica napus*, *Plant Mol. Biol.* **37**, 763-772.
- Samach, A., Onouchi, H., Gold, S.E., Ditta, G.S., Schwarz-Sommer, Z., Yanofsky, M.F. and Coupland, G. (2000) Distinct roles of *CONSTANS* target genes in reproductive development of *Arabidopsis*, *Science* **288**, 1613-1616.
- Samach, A. and Coupland, G. (2000) Time measurement and the control of flowering in plants, *BioEssays* **22**, 38-47.

- Sarma, R.N., Gill, B.S., Sasaki, T., Galiba, G., Sutka, J., Laurie, D.A. and Snape, J.W. (1998) Comparative mapping of the wheat chromosome 5A *Vrn-A1* region with rice and its relationship to QTL for flowering time, *Theor. Appl. Genet.* **97**, 103-109.
- Sasaki, T. and Burr, B. (2000) International Rice Genome Sequencing Project: The effort to completely sequence the rice genome, *Curr. Opin. Plant Biol.* **3**, 138-141.
- Schmidt, R. (2000) Synteny: recent advances and future prospects, *Curr. Opin. Plant Biol.* **3**, 97-102.
- Shahla, A. and Tsuchiya, T. (1990) Genetic-analysis in 6 telotrisomic lines in barley (*Hordeum vulgare* L.), *J. Hered.* **81**, 127-130.
- Sheldon, C.C., Burn, J.E., Perez, P.P., Metzger, J., Edwards, J.A., Peacock, W.J. and Dennis, E.S. (1999) The *FLF* MADS box gene: A repressor of flowering in *Arabidopsis* regulated by vernalization and methylation, *Plant Cell* **11**, 445-458.
- Sheldon, C.C., Rouse, D.T., Finnegan, E.J., Peacock, W.J. and Dennis, E.S. (2000) The molecular basis of vernalization: The central role of *FLOWERING LOCUS C (FLC)*, *Proc. Natl. Acad. Sci. USA* **97**, 3753-3758.
- Shirasu, K., Lahaye, T., Tan, M.W., Zhou, F.S., Azevedo, C. and Schulze-Lefert, P. (1999) A novel class of eukaryotic zinc-binding proteins is required for disease resistance signaling in barley and development in *C-elegans*, *Cell* **99**, 355-366.
- Simpson, G.G., Gendall, A.R. and Dean, C. (1999) When to switch to flowering, *Annu. Rev. Cell Dev. Biol.* **15**, 519-550.
- Strayer, C., Oyama, T., Schultz, T.F., Raman, R., Somers, D.E., Más, P., Panda, S., Kreps, J.A. and Kay S.A. (2000) Cloning of the *Arabidopsis* clock gene *TOC1*, an autoregulatory response regulator homolog, *Science* **289**, 768-771.
- Stelmakh, A.F. (1998) Genetic systems regulating flowering response in wheat, *Euphytica* **100**, 359-369.
- Stracke, S. and Börner, A. (1998) Molecular mapping of the photoperiod response gene *ea*⁷ in barley, *Theor. Appl. Genet.* **97**, 797-800.
- Sugano, S., Andronis, C., Ong, M.S., Green, R.M. and Tobin, E.M. (1999) The protein kinase *CK2* is involved in regulation of circadian rhythms in *Arabidopsis*, *Proc. Natl. Acad. Sci. USA* **96**, 12362-12366.
- Swarup, K., Alonso-Blanco, C., Lynn, J.R., Michaels, S.D., Amasino, R.M., Koornneef, M. and Millar, A.J. (1999) Natural allelic variation identifies new genes in the *Arabidopsis* circadian system, *Plant J.* **20**, 67-77.
- Takahashi, R. and Yasuda, S. (1971) Genetics of earliness and growth habit in barley, in: R.A. Nilan (ed.), *Proc. 2nd Int. Barley Genet. Symp.*, Washington State University Press, pp. 388-408.
- Takahashi, Y., Shomura, A., Sasaki, T. and Yano, M. (2001) *Hd6*, a rice quantitative trait locus involved in photoperiodic sensitivity, encodes the α subunit of protein kinase CK2, *Proc. Natl. Acad. Sci. USA* **98**, 7922-7927.
- Tanksley, S.D. and Nelson, J.C. (1996) Advanced backcross QTL analysis: A method for the simultaneous discovery and transfer of valuable QTLs from unadapted germplasm into elite breeding lines, *Theor. Appl. Genet.* **92**, 191-203.
- Taylor, S.A. and Murfet, I.C. (1996) Flowering in *Pisum*: Identification of a new *ppd* allele and its physiological action as revealed by grafting, *Physiol. Plant.* **97**, 719-723.
- Thomas, B. and Vince-Prue, D. (1997) *Photoperiodism in plants*, Academic Press, London.

- Thomas, W.T.B., Powell, W., Waugh, R., Chalmers, K.J., Barua, U.M., Jack, P., Lea, V., Forster, B.P., Swanston, J.S., Ellis, R.P., Hanson, P.R., Lance, R.C.M. (1995) Detection of quantitative trait loci for agronomic, yield, grain and disease characters in spring barley (*Hordeum vulgare* L.), *Theor. Appl. Genet.* **91**, 1037-1047.
- Tinker NA, Mather DE, Rosnagel BG, Kasha KJ, Kleinhofs A, Hayes PM, Falk DE, Ferguson T, Shugar LP, Legge WG, Irvine RB, Choo TM, Briggs KG, Ullrich SE, Franckowiak JD, Blake TK, Graf RJ, Dofing SM, Saghai Maroof MA, Scoles GJ, Hoffman D, Dahleen LS, Kilian A, Chen F, Biyashev RM, Kudrna DA, Steffenson BJ (1996). Regions of the genome that affect agronomic performance in two-row barley, *Crop Sci.* **36**, 1053-1062.
- Tranquilli, G. and Dubcovsky, J. (2000) Epistatic interaction between vernalization genes *Vrn-A^m1* and *Vrn-A^m2* in diploid wheat, *J. Hered.* **91**, 304-306.
- Vladutu, C., McLaughlin, J. and Phillips, R.L. (1999) Fine mapping and characterization of linked quantitative trait loci involved in the transition of the maize apical meristem from vegetative to generative structures, *Genetics* **153**, 993-1007.
- Van Deynze AE, Nelson JC, Yglesias ES, Harrington SE, Braga DP, McCouch SR, Sorrells ME (1995) Comparative mapping in grasses. Wheat relationships, *Mol. Gen. Genet.* **248**, 744-754.
- Vos, P., Hogers, R., Bleeker, M., Reijans, M., van de Lee, T., Hornes, M., Frijters, A., Pot, J., Peleman, J., Kuiper, M. and Zabeau, M. (1995) AFLP: a new technique for DNA fingerprinting, *Nucl. Acids Res.* **23**, 4407-4414.
- Walbot, V. (2000) Saturation mutagenesis using maize transposons, *Curr. Opin. Plant Biol.* **3**, 103-107.
- Weller, J.L., Reid, J.B., Taylor, S.A. and Murfet, I.C. (1997a) The genetic control of flowering in pea. *Trends Plant Sci.* **2**, 412-418.
- Weller, J.L., Murfet, I.C. and Reid, J.B. (1997b) Pea mutants with reduced sensitivity to far-red light define an important role for phytochrome A in day-length detection, *Plant Physiol.* **114**, 1225-1236.
- Worland, A.J. (1996) The influence of flowering time genes on environmental adaptability in European wheats, *Euphytica* **89**, 49-57.
- Worland, A.J. and Law, C.N. (1986) Genetic analysis of chromosome 2D of wheat. I. The location of genes affecting height, day-length insensitivity, hybrid dwarfism and yellow-rust resistance, *Z. Pflanzenzüchtg* **96**, 331-345.
- Worland, A.J., Börner, A., Korzun, V., Li, W.M., Petrovic, S. and Sayers, E.J. (1998) The influence of photoperiod genes to the adaptability of European winter wheats, *Euphytica* **100**, 385-394.
- Yamamoto, T., Kuboki, Y., Lin, S.Y., Sasaki, T. and Yano, M. (1999) Fine mapping of quantitative trait loci *Hd-1*, *Hd-2* and *Hd-3*, controlling heading date of rice, as single Mendelian factors, *Theor. Appl. Genet.* **97**, 37-44.
- Yamamoto, T., Lin, H.X., Sasaki, T. and Yano, M. (2000) Identification of heading date quantitative trait locus *Hd6* and characterization of its epistatic interactions with *Hd2* in rice using advanced backcross progeny, *Genetics* **154**, 885-891.
- Yano, M., Katayose, Y., Ashikara, M., Yamanouchi, U., Monna, L., Fuse, T., Baba, T., Yamamoto, K., Umehara, Y., Nagamura, Y. and Sasaki, T. (2000) *Hd1*, a major photoperiod sensitivity quantitative trait locus in rice, is closely related to the *Arabidopsis* flowering time gene *CONSTANS*, *Plant Cell* **12**, 2473-2483.

- Yano, M., Harushima, Y., Nagamura, Y., Kurata, N., Minobe, Y. and Sasaki, T. (1997) Identification of quantitative trait loci controlling heading date in rice using a high-density linkage map, *Theor. Appl. Genet.* **95**, 1025-1032.
- Zhu, H., Gilchrist, L., Hayes, P., Kleinhofs, A., Kudrna, D., Liu, Z., Prom, L., Steffenson, B., Toojinda, T. and Vivar, H. (1999) Does function follow form? Principal QTLs for *Fusarium* head blight (FHB) resistance are coincident with QTLs for inflorescence traits and plant height in a doubled-haploid population of barley, *Theor. Appl. Genet.* **99**, 1221-1232.

Web addresses

TAIR (<i>Arabidopsis</i>)	http://www.arabidopsis.org/
CerealsDB (Triticeae)	http://jiiio5.jic.bbsrc.ac.uk:8000/
Graingenes (Triticeae)	http://wheat.pw.usda.gov/
MaizeDB	http://www.agron.missouri.edu/
Rice (Japan)	http://bank.dna.affrc.go.jp/
RiceDB (USA)	http://ars-genome.cornell.edu/rice/
Soybase (soybean)	http://129.186.26.94/
SorghumDB	http://algodon.tamu.edu/sorghumdb.html
SolGenes (Solanaceae)	http://ars-genome.cornell.edu/solgenes/welcome.html

MOLECULAR MARKERS FOR THE GENETIC ANALYSIS OF APOMIXIS

M.J. ASINS, M.R. GARCIA, C. RUIZ and E.A. CARBONELL
I.V.I.A., Apdo. Oficial; 46113 Moncada (Valencia), Spain

1. INTRODUCTION

Most flowering plants reproduce sexually through seeds. A zygote is formed by fusion of reduced female and male gametes (amphimixis) and develops into an embryo. However, some angiosperms commonly reproduce through seeds having an embryo, which is formed without reduction of the chromosome number nor fertilisation, by a process called apomixis (Nogler, 1984).

To denote the multiplicity and continuity of reproductive systems in flowering plants, Fryxell (1957) represented them schematically as a triangle. The three vertices correspond to absolute allogamy, absolute autogamy and obligate apomixis where species rarely position themselves somewhere on the sides of the triangle. In general, few species are obligate apomicts. Most apomictic species are facultative, allowing a variable degree of sexual reproduction. In fact, facultative apomixis seems to confer versatility, adaptedness and evolutive capacity to some species as *Poa pratensis* L. (Mazzucato *et al.*, 1996). Clausen (1961) explained this evolutionary adaptability by the dual ability of facultative apomictics to occasionally side-step sexual reproduction and to multiply the successful combinations asexually.

Apomictic processes have been observed in more than 400 plant species belonging to 35 families, the most well represented including the Gramineae, Compositae, Rosaceae and Rutaceae (Hanna and Bashaw, 1987; Carman, 1997). Apomixis occurs in several species of agricultural value. Among the grasses it prevails in polyploid species of *Paspalum*, *Panicum*, *Poa*, *Bothriochloa*, *Dichanthium*, *Eragrostis*, *Pennisetum* and *Cenchrus* (Barshaw, 1975). Other important genera include *Citrus* and *Rubus*, and crops as diverse as sugar beet, apple, pearl millet, wheat and maize rank apomictic species among their wild relatives.

A number of known mutations cause autonomous diploid endosperm development and even haploid embryo structures were observed without a preceding fertilisation step (in Luo *et al.*, 1999); other mutations affect genotypic transmission, for example, by eliminating recombination or the first meiotic nuclear division. However, no gametophytic or sporophytic mutants have been recovered to date in sexual species, suggesting they require gains of function (Vielle Calzada *et al.*, 1996), not just the knock out of gene(s) controlling the sexual reproduction process.

Given that seeds resulting from apomictic reproduction contain embryos whose genetic constitution is identical to that of the female parent, apomixis is seen as a potentially powerful breeding tool to fix heterosis. Thus, apomixis would allow fixing the genotype of a superior variety, including hybrid cultivars, and hence seed could be produced for many generations without loss of vigour or genotypic segregation, enabling a significant reduction in hybrid seed production costs. Therefore, most research to date has focused on introgressing the trait of apomixis by traditional breeding into agricultural crops of economical importance, such as wheat and maize from wild, often very distant relatives. However, this is a slow process because it involves wild relatives, generally with different ploidy level, small progeny sizes and a trait that is laborious to evaluate (progeny tests). As Koltunow *et al.* (1995) points out, these problems can be overcome if molecular knowledge of the genes involved in initiating and controlling apomixis were available because they could be transferred to the crop of interest by molecular transformation. Co-segregational analysis of molecular markers and traits is a powerful methodology to study apomixis genetic control: number of genomic regions involved, their individual contribution and the estimation of gene effects and interactions. The analysis allows the early selection of putative apomictic progeny by genotyping molecular markers linked to the apomixis genes. This is specially valuable in cases such as apomixis where the evaluation of the trait is laborious and time consuming. Moreover, if genes of known biochemical function are also included among the markers it is possible to check their role as candidate genes explaining some QTLs involved in the trait.

The aim of the present chapter is to review how much light the analysis of molecular markers has provided about the genetic control of apomixis and what is still left unclear.

2. TYPES OF APOMIXIS

Given that the genetic control inferred by marker analysis of apomixis must fit observations at biochemical, cytological and seed levels, the

apomictic process has to be first clarified. There is a sequence of events that take place in the ovule and results in a seed (after Koltunow, 1993): megaspore mother cell differentiation from the nucellus, megaspore production by meiosis, megaspore selection, embryo sac development by mitotic divisions, embryo sac maturation, double fertilisation, and endosperm and embryo formation. Apomictic processes are usually classified into three types differing in the time at which it is initiated during the ovule development relative to the normal sexual pathway: diplospory, apospory and adventitious (nucellar) embryony. Diplospory and apospory are initiated early in ovule development; at the time of megaspore mother cell differentiation in the case of diplospory and after mother cell differentiation in apospory, resulting both in the formation of an unreduced megagametophytic structure without meiotic reduction. Diplospory and apospory are, therefore, commonly referred to as gametophytic apomixis. In diplospory, the megaspore mother cell switches from a sexual to an apomictic pathway resulting in a parthenogenetic embryo from an unreduced embryo sac. On the other hand, aposporous embryo sacs form from additional nucellar cells following megaspore mother cell differentiation. Sexual and aposporous processes can coexist in one ovule, which by definition is not possible for diplosporous apomicts but this does not preclude sexual and apomictic processes coexisting amongst ovules of diplosporous apomicts.

By contrast to gametophytic apomixis, adventitious embryony is initiated late in ovule development at the nucellus, generally, after embryo sac maturation. In adventitious embryony (also named sporophytic embryony), embryos develop from cells in tissues external to a sexual embryo sac, the nucellus or the inner integument (Lakshmanan and Ambegaokar, 1984). The nucellar cells destined to become embryos are morphologically distinguished from the surrounding nucellar cells by their large nuclei and dense cytoplasm, a morphology reminiscent of that of aposporous initial cells (Koltunow, 1993). Most species with adventitious embryony like *Citrus* require endosperm generated by pseudogamy for seed set (Asker and Jerling, 1992). Thus, the initiation of nucellar embryo development in *Citrus* seems to be independent of fertilisation, but endosperm formed in the sexual embryo sac is essential for its nutrition and ulterior development (Koltunow, 1993).

The nucellar tissue of the ovule is the source of female (reduced or unreduced) gametophyte and also the source of apomictic embryos, with few exceptions (those originated from the integuments). In plants that reproduce sexually, the initiation of Polygonum-type embryo sac development is restricted to a single cell, which differentiates from the nucellus. By contrast,

the nucellus of apomictic species has a greater developmental plasticity than that of sexual reproducing plants.

Most plants with gametophytic apomixis are polyploid; however, genera with adventitious embryony are commonly diploid (Asker and Jerling, 1992; Carman, 1997).

3. GENETIC CONTROL OF GAMETOPHYTIC APOMIXIS

3.1. Diplospory

Molecular markers have been used for the genetic analysis of diplospory in the *Tripsacum* ssp. ($x=18$), the closest wild apomict relative to *Zea mays* L. ($x=10$), one of the most important cereal crop in the world. Taking advantage of the numerous genome similarities to the maize genome, Leblanc *et al.* (1995a) chose 93 DNA clones detecting loci throughout the maize genome (at 20-40 cM intervals) to hybridise with DNA from parentals *T. dactyloides* ($2n=4x=72$), a maize hybrid and bulks from diplosporous and sexual progenies. Three linked RFLP markers located at the distal end of the long arm of chromosome 6 co-segregated with diplospory. Kindiger *et al.* (1996b), combining cytological and molecular analysis of 23 individuals reported that the long arm of *Tripsacum* chromosome 16 carries the gene(s) for apomictic reproduction because all apomictic individuals (14) carry this chromosome arm and sexuals (9) do not. Noteworthy, all apomictic individuals carry at least 7 additional *Tripsacum* chromosomes. Comparison of the maize and *Tripsacum* molecular maps indicates that the distal region of the long arm of *Tripsacum* chromosome 16 has colinearity with the distal region of maize chromosome 6. Both homeologous chromosomes possess a nucleolar organising region that is identified by a secondary constriction.

A key element for the genetic analysis of a character is the way it has been evaluated. Moreover, analysis of co-segregation of markers and the trait needs a fine evaluation of the trait in the progeny. If the trait is evaluated not by itself but rather by a correlated trait, then caution must be taken about the conclusions obtained driven for the original trait. Leblanc *et al.* (1995a) classified the progeny categorically into apomictic or sexual hybrids. Fifty-two hybrids ($2n=46$) were evaluated by analysis of callose deposition during megasporogenesis because the determination of the mode of reproduction of each hybrid through progeny-tests (progeny heterogeneity) could not be carried out. In previous studies, Leblanc *et al.* (1995b) found that non-reduction of chromosomes (meiosis failure) is

strongly associated with an absence of callose deposition in megasporocyte cell wall, whereas meiotic megasporocytes and their derivatives are surrounded by callose. These cytological studies undoubtedly reveal the mechanism of apomixis (meiotic diplospory) but progeny-tests are still required to determine the degree of apomixis of hybrids. In their study, the trait (degree of apomixis) was not analysed by molecular markers but a correlated trait due to the fertility limitations originated from the cross. Kindiger *et al.* (1996a) evaluated apomixis by the lack of variation within families using cytological, isozyme and RAPD analysis; nevertheless, apomictic individuals were considered 100 % apomicts (obligate apomicts) or 100 % sexuals in the genetic analysis of the character (Kindiger *et al.*, 1996b).

3.2. Apospory

The best known example of genetic analysis of apospory using molecular markers has been reported by Ozias-Akins *et al.* (1998). The mapping population (397 individuals) was produced from a cross between an induced tetraploid line of *Pennisetum glaucum* ($2n=6x=54$) and *P. squamulatum* ($2n=6x=54$), an obligate apomict as the male parent. Apospory was evaluated as the presence of four-nucleate embryo sacs (indicative of aposporous embryo sacs), from microscopic examination of 20 cleared ovules from each individual. The mode of reproduction was confirmed later by examining progeny produced after pollination of more than 100 hybrids with pollen from a tetraploid pearl millet genotype homozygous for a dominant “red” marker gene, i. e., progeny test was used to evaluate the trait. Quite surprisingly, all this information was not included in the molecular marker analysis but each progeny was simply categorised only as sexual or aposporous (at any degree). Bulk segregant analysis was carried out using only segregating alleles (RAPDs) from heterozygous *P. squamulatum* and ignoring pearl millet alleles. Using this strategy they were able to obtain 12 PCR-base markers that co-segregated with apospory (except for one hybrid) in coupling phase. Hybridisation analysis using 6 of the 12 markers as probes showed that hybridisation signals were observed only in the apomictic individuals for 4 of the probes, suggesting that hybrids were hemizygous at these loci. Due to probabilistic reasons, all 12 markers must belong to the same linkage group or chromosome defining what authors called apospory-specific genomic region (ASGR). Their hypothesis is that ASGR represents a complex locus where gene silencing might be involved as part of the apomixis mechanism (silencing or repression of genes required for normal sexual reproduction). Certainly, a silencing locus would fit the dominance of the trait, its incomplete penetrance in facultative apomicts and

its potential for suppression by other loci. This suppression of the apomixis gene would explain the presence of ASGR in one their sexual hybrids. Evidence for suppression of apomixis has been observed in *Cenchrus ciliaris* (Taliaferro and Bashaw, 1966) and *Citrus* (Tisserat *et al.*, 1978), which will be discussed later in the part of nucellar embryony.

3.3. Evolutionary Relationships of Gametophytic Apomixis Genes

Roche *et al.* (1999) have recently shown that ASGR is highly conserved between *Pennisetum squamulatum* and *Cenchrus ciliaris*. Six out of the 12 molecular markers linked to apomixis in *Pennisetum* are also linked to apomixis in *C. ciliaris*.

The most recent data from molecular marker analysis of gametophytic apomixis in species like *Tripsacum dactyloides* (Leblanc *et al.*, 1995a), *C. ciliaris*, *P. squamulatum* (Roche *et al.*, 1999) and *Brachiaria* (Pessino *et al.*, 1997), strongly supports dominant inheritance of a single-gene block trait. All of these species are polyploid (4x or 6x), without strict bivalent pairing and homeologous pairing of chromosome bearing the apomixis gene(s) is unknown.

Since a common segregation pattern for all 4 species was often a ratio less than the predicted 1 apomict to 1 sexual for a single dose of the apomixis dominant allele, a gametic recessive lethal factor linked to the apomixis locus has been always hypothesised. Furthermore, polyploidy, due to occasional fusion of unreduced male gametes (which would be the only viable male gametes carrying the apomixis trait) with unreduced female gametes, should be a natural outcome and could explain the predominance of polyploidy in gametophytic apomicts. A different mechanism has been suggested by Bicknell *et al* (2000) in *Hieracium* spp where the formation of diploid apomicts is discouraged through selection against diploid hybrids, acting after fertilization, rather than through gamete-level selection. In fact, most progeny of the triploid apomict is also triploid, not only when the sexual female parent is diploid but also when it is tetraploid, suggesting the action of an important selection pressure on the ploidy level of the progeny itself.

Since one single locus or block of genes is always reported to control gametophytic apomixis, a question arises: do evolutionary related genes control apospory and diplospory? There is an indirect evidence suggesting that the answer is no. This evidence lies on the following observation: there are two wild relatives of maize, one showing apospory (*Brachiaria brizantha*) and the other diplospory (*Tripsacum dactyloides*). While

apospory locus map near RFLPs that correspond to a duplicated set of loci in maize chromosomes 1 and 5, the RFLP markers linked to diplospory correspond to clones that map at maize chromosome 6.

4. GENETIC CONTROL OF ADVENTITIOUS EMBRYONY

The nucellar form of adventitious embryony is most common and will be discussed here because much insight into this process has been gained in citrus during many years including cytological, inheritance (genetic and non-genetic factors) and recently, marker analysis. The reason for so much interest is clear. Nucellar embryony has very important consequences for breeding and culture of citrus fruits. Budding onto seedling rootstocks almost universally propagates citrus. Therefore, propagation of citrus rootstocks depends upon the production of clonal plants from nucellar seedlings. Uniformity of rootstock genotype is essential for reliable plant performance following budding and orchard establishment. Most rootstock cultivars presently grown are polyembryonic, producing seeds that contain both nucellar and zygotic embryos. Therefore, one of the most important traits within the breeding programs for citrus rootstock is apomictic reproduction by seed, i.e. that the new genotype yields the least number of zygotic seedlings.

The presence of extra (nucellar) embryos in addition to that from sexual origin leads to a closely related term, polyembryony. Polyembryony is the development of two or more embryos in one seed. This is why a polyembryonic variety is commonly considered apomictic (yielding apomictic, nucellar, seedlings) and a monoembryonic variety, sexual or zygotic seedlings. However, it is known that extra embryos are occasionally sexual, produced either from fission of one fecundated egg or from two or more functional embryo sacs in a single ovule or simple budding processes from the growing zygotic embryo, yielding clones of the zygote (Bacchi 1943, Cameron and Garber 1968).

Segregation analysis of polyembryony in terms of proportion of offspring has been previously reported (Parlevliet and Cameron 1959; Iwamasa *et al.*, 1967; Cameron and Soost 1979). Three kind of crosses are usually studied, monoembryonic x monoembryonic, monoembryonic x polyembryonic and polyembryonic x polyembryonic, including several monoembryonic and polyembryonic cultivars as parentals. Embryo counts from 50-100 seeds per hybrid allowed these authors an accurate estimation of the degree of polyembryony; however, hybrids having 1-6 % polyembryonic seed were

classed as monoembryonic; those showing more than 20 % polyembryonic seeds were classed as polyembryonic and those between 6 and 20 omitted in the studies. Taken into account this simplification and averaging over progenies within cross type, the reported results mostly fit statistically the presence of one principal dominant gene controlling the occurrence of polyembryony and the homozygous condition of a recessive allele determining monoembryony. Nevertheless, in some crosses, progenies varied widely from 1:1 and 3:1 ratios suggesting more genes involved. In fact, as pointed out by the authors, a variable degree of polyembryony among polyembryonic hybrids is found, implying the presence of minor genes affecting this degree.

Other results supporting the continuous nature of the polyembryony trait involve environmental factors such as the year, pollen origin, tree vigour, eco-geographical location of the tree, the north-south position of the fruit within the tree, the age of the plant and the rootstock (grafting a polyembryonic cultivar on a monoembryonic rootstock reduces the rate of nucellar embryony) (Furusato 1954, Parlevliet and Cameron 1959; Minessey 1953; Garcia *et al.*, 1999).

Garcia *et al.* (1999) reported the first genetic study of polyembryony and apomixis using molecular marker analysis in a cross between two diploid ($2n=18$) polyembryonic species (*Citrus volkameriana* and *Poncirus trifoliata*), differing in the degree of polyembryony (CxP population), produced by Dr. J.B. Forner at IVIA 20 years ago.

Apomictic reproduction was studied by genotyping for isozymatic loci (6 enzymatic systems) twenty-five random seedlings derived from seeds of each individual fruit- yielding tree (50 out of 80 trees). Enzymatic loci at which progeny trees were heterozygous were chosen to estimate the percentage of zygotic seedlings per tree. Seedlings that were heterozygous at all those loci were considered nucellar (sexual seedlings always derived from self-pollination of the mother plants). For some trees, this evaluation was carried out for two consecutive years. This progeny-test allowed the estimation of the degree of apomixis. Nevertheless, this estimation can not be considered very accurate for several reasons. First, apomixis is evaluated as a percentage obtained from a 25 plant-sample, therefore there is a 4% intrinsic error. Secondly, seedlings arising from sexual embryos may present the same genotype than the mother hybrid at, say, 5 loci, although the probability of this event is low ($p=(1/2)^5=0.03125$).

Based on the fact that monogenic inheritance had been suggested by several authors (gametophytic apomixis) and on the segregation ratio found in this study, bulked segregant analysis was the strategy used for searching the chromosomal position. Segregation of a total of 69 markers (8 microsatellites, 43 RAPDs, 13 RFLPs, 1 CAPS and 4 isozymatic loci) was

analysed in the whole CxP population. *C. volkameriana* was heterozygous at 45 marker loci and *P. trifoliata* was at 38 marker loci. In *C. volkameriana* distorted segregation was found for 22.22% of the markers; in *P. trifoliata*, for 47.37 %. Two markers showed distorted segregation in both species. Distorted segregation at the genotypic level could always be explained just by gametic selection.

Analysis of quantitative trait loci (QTL) revealed the presence of 6 genomic positions (two in *P. trifoliata* and four in *C. volkameriana*) contributing individually up to a 24 % of total variation for apomixis. Within the same species, QTLs with positive and negative allele effects are present, even in the same linkage group. One of the markers associated to apomixis (*Apo2*) is also associated to embryony type. Therefore the genetic control of apomictic reproduction found in citrus for nucellar embryony is quite complex compared to what had been reported for gametophytic apomixis.

Before trying to interpret biologically these results on the genetic control of nucellar embryony, it is important to discuss the limitations of this study. As pointed out by the authors, main limitations of the QTL analysis arise from progeny size (only 50 out of 80–tree progeny yielded fruit), only two parents were studied, the number of molecular markers, the distorted segregation of some of them (specially for the male parent), the way apomixis was evaluated and the lack of normality of the trait.

At present moment, we cannot improve the progeny size and no other cross is available yet to evaluate apomixis but we can play around with the other factors.

García *et al.* (1999) found that some hybrids classified as monoembryonic showed a certain level, although a low one, of apomixis. We have found that monoembryony was not absolute but, for some rechecked hybrids, a small proportion of seeds (1 out of 20 seeds, in some cases) presented 2 or 3 embryos; i. e. monoembryony/polyembryony has to be considered as a continuous trait and expressed in terms of percentage. Other, not-checkable possibility has been brought about by Dr. M. Roose (personal communication); considering a hybrid that exclusively yields sexual seedlings (a monoembryonic one), the probability of having, at least, 1 among 25 seedlings genetically identical to the mother hybrid plant is 0.548 (Table 1). Therefore, this could be the case of some monoembryonic (obligate sexual) hybrids with 4% “putative apomictic” seedlings.

To study the possible error in the previous QTL analysis of apomixis we are now presenting a comparison of QTLs detected using apomixis both as originally evaluated and re-coded by changing the low % apomixis of monoembryonic hybrids to 0% apomixis. Do some QTLs disappear? Do we get a simpler genetic control?

This new analysis of QTLs has been carried out now using the Kruskal-Wallis statistical methodology, suitable for non-normal distributed traits, and increasing the number of polymorphic markers, 73 in *P. trifoliata* and 97 in *C. volkameriana*. New markers are 43 SSR loci (Figure 1) and 11 IRAPs (unpublished data).

Table 1. N is the number of heterozygous progeny at 5 loci derived by self-pollination of the heterozygous mother plant. P, the probability of having N of this kind within a progeny size of 25 and ΣP is the cumulative probability. Last column represents the decreasing probability of having at least N heterozygous progenies at 5 loci in 25 when N increases.

CLASS (N)	P	ΣP	P(at least N in 25)
0	0.452	0.452	
1	0.365	0.817	0.548
2	0.141	0.958	0.183
3	0.035	0.993	0.042
4	0.006	0.999	0.007
5	0.001	1.000	0.001

The results are presented in Table 2. Comparing to QTLs reported by Garcia *et al.* (1999) only *Apo 5* on *P trifoliata* linkage group P4 disappears and, in fact, this QTL was not detected in the previous study using the t-test mean comparison method. On the other hand, as expected from the fact of analysing more markers, two new *Apo* QTLs are detected, in *C. volkameriana* linkage maps (now named *Apo5* and *Apo7*). Therefore, the picture does not seem to change much.

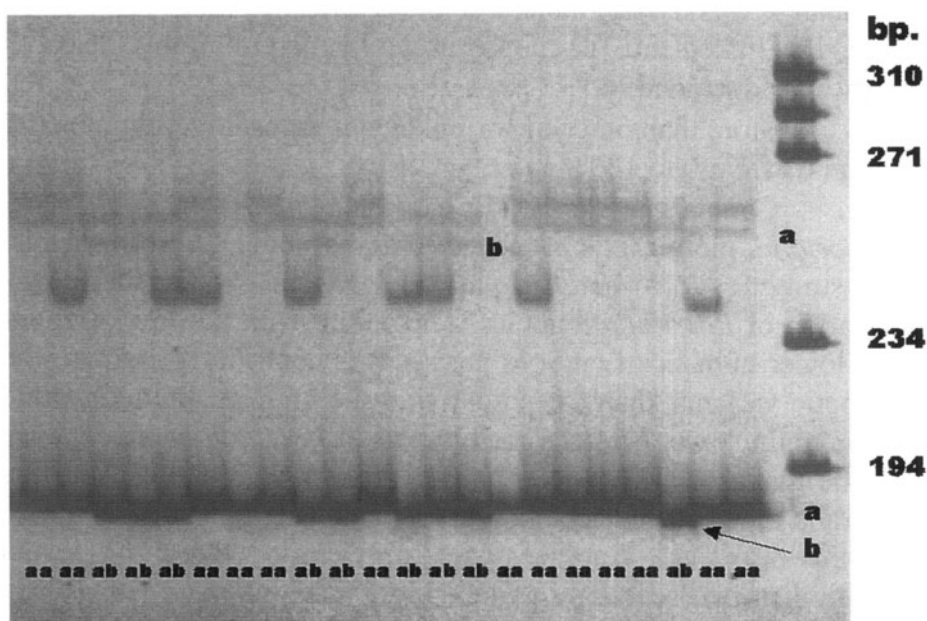


Figure 1.- Silver stained polyacrylamide gel showing segregation at microsatellite marker CR5-260, linked to *Apo7*.

Table 2. Detection of apomixis QTLs using the actual apomixis values and the re-coded ones (in italics) for *C. volkameriana* and *P. trifoliata* segregating alleles. L: Linkage group; V for *C. volkameriana* and P for *P. trifoliata*. SL: Significance level. *: 10%; **: 5%; ***: 1%; *****: 0.05%

L	Marker	SL	QTL	Means	SL	Means
V8	OPG13075	**		63.6:35.0	*	60.4:32.0
	TAA52	**	<i>Apo6</i>	32.5:62.5	**	29.4:59.5
	OPG19200	*		37.8:60.9		
V8	OPE041000	**	<i>Apo1</i>	34.8:66.9	**	29.9:66.1
V4	CR3	**	<i>Apo5</i>	35.6:58.7	*	32.9:55.2
V7	ERS	**	<i>Apo3</i>	18.3:55.4	**	13.8:52.9
V7	CR7	**	<i>Apo7</i>	23.1:58.8	**	18.9:56.1
	CR5-260	**		59.4:26.1	**	56.6:22.3
V13	TAA15	***	<i>Apo2</i>	69.6:25.2	*****	69.6:18.8
P3	Egp47	**	<i>Apo3</i>	82.6:38.3	*	77.6:35.9

Comparing the QTLs detected with the actual apomixis values and those detected when data are re-coded (monoembryonic hybrids with low apomixis % changed to 0% apomixis) few changes occur. The most

important change was the large increase of significance reached by TAA 15, the *Apo2* QTL, linked also to monoembryony/polyembryony. Then, from both analyses is concluded:

- More than one gene or linkage group is governing the degree of apomixis from 0 to 100%.
- Comparing the QTLs detected in *P. trifoliata* and *C. volkameriana*, more QTLs are detected and at higher significance levels when studying *C. volkameriana* alleles than when studying the segregation of *P. trifoliata* alleles. The parent with the lowest apomixis value (lower number of embryo per seed) is contributing much more to the genetic variance than the parent with the highest apomixis value. In addition to the lower heterozygosity value of *P. trifoliata*, an explanation might be that *Apo* QTL alleles at *P. trifoliata* have similar substitution effects and the progeny size does not provide enough statistical power to detect this difference. Another related problem is the high proportion of loci with distorted segregation, including those linked to *Apo 4*. It is not the first time a gametic lethal factor is found linked to an apomixis gene.
- The strongest effect on apomixis seems to correspond to *Apo 2* (a *C. volkameriana* QTL), also related to monoembryony / polyembryony. Is this QTL a major gene?

4.1. An Hypothesis on the Regulation of Apomixis in Citrus

There is a possible connection among these results and previous observations referred earlier in this chapter. Esan (1973) showed that seeds of naturally polyembryonic members of the Rutaceae were apparently low in a factor that inhibited nucellar embryogenesis. From experiments using seeds of mono and polyembryonic, he noted that an inhibitory factor could be transmitted from monoembryonic to polyembryonic tissue through seed grafts as well as through the nutrient medium. Tisserat and Murashige (1977) discovered subsequently that the factor was comprised of volatile and non-volatile substances, and possibly included ethylene, abscisic acid, auxin, gibberellin and ethanol. The Esan data implies that sexual citrus plants are sexual because they contain high levels of a factor that repress embryogenesis of nucellar cells while apomictic varieties contain less of the factor and this somehow allows nucellar cells to undergo direct embryogenesis. Based upon all these observations and present results of genetic analysis we hypothesise the following model of nucellar embryony in Citrus:

All Citrus species might be potentially polyembryonic given that nucellar plants of monoembryonic citrus cultivars have been obtained long time ago

by *in vitro* culture of nucellus (Tisserat *et al.*, 1977); i. e., it is an apomictic genus where some species are monoembryonic because they have a system to repress the development of embryos from nucellus. This would explain why most QTLs are found when the segregation of the least polyembryonic parent is studied.

The repression would happen when a certain “complex substance” reaches a given level. Why sexual embryo development is not inhibited too? Guerin *et al.* (2000) have recently shown, by *in-situ* hybridization, that megaspore mother cells and ovule cells where aposporous initial cells differentiate, can be differentially regulated from each other and from the other ovule cells. Therefore, aposporous cell initials may have the receptors for the “inhibitory substance” whilst megaspore mother cells do not.

The “complex substance” would include some plant hormones such as ethylene, which inhibit nucellar embryogenesis *in vitro* (Tisserat and Murashige, 1977). *Apo 3* from *C. volkameriana* has shown an effect decreasing apomixis and it is important to point out that this QTL is revealed when a DNA clone corresponding to an ethylene receptor gene is used as marker (Garcia *et al.*, 1999). Therefore ethylene, among other substances could be repressing the development of nucellar embryos through its union with an ethylene receptor protein of nucellar cells. A related observation might be the higher mean number of embryos per seed in seeds from north side fruits of a tree than in those from the south side (Furusato, 1954; Parlevliet and Cameron, 1959). The amount of ethylene, abscisic acid or auxin in the fruits of a tree might not be uniformly distributed but northern side fruits might produce fewer amounts as average because temperature or sunlight differences. This would explain why seeds from fruits from the northern side of the tree have as average more embryos per seed than those from the southern side.

Among the QTLs found, *Apo2*, linked to TAA15, show a special relevance and present negative effects on apomixis. *Apo 2* might control a key step in the synthesis of the inhibitory “complex substance” during fruit maturation. The other QTLs and probably others not yet detected, would be involved in the synthesis, degradation and signal transduction of the compounds that constitute the “complex substance” hypothesised by Esan.

5. CONCLUDING REMARKS AND PERSPECTIVES

The genetic analysis of apomixis by molecular markers has provided useful information about this agronomically important mode of reproduction, nevertheless the puzzle is far from being complete yet. What do we have until now?

- There are three main developmental processes driving to apomixis that are distinguished by cytological analysis. Two of these processes are gametophytic (diplosporous and aposporous apomixis) and one, sporophytic (adventitious embryony).

- Data from marker analysis of gametophytic apomixis in species like *Tripsacum dactyloides* (Leblanc *et al.*, 1995a), *C. ciliaris*, *P. squamulatum* (Roche *et al.*, 1999) and *Brachiaria* (Pessino *et al.*, 1997), strongly support dominant inheritance of a single-gene block trait. All of these species are polyploid (4x or 6x), without strict bivalent pairing and homeologous pairing of chromosome unknown.

- There are several markers that co-segregate with aposporous apomixis and must define a specific genomic region only present in apomictic genotypes and called apospory-specific genomic region (ASGR). ASGR is highly conserved between *Pennisetum squamulatum* and *Cenchrus ciliaris*, two phylogenetically related apomictic species

- Most inheritance studies have considered apomixis as a categorical trait not as a continuous one.

- More than one gene or linkage group is governing the degree of sporophytic apomixis from 0 to 100% in *Citrus*.

- The strongest effect on this study of apomixis seems to correspond to *Apo 2* (a *C. volkameriana* QTL with negative gene effects), also related to monoembryony/polyembryony.

- One genomic region seems to have important effects on the degree of both gametophytic and sporophytic apomixis. These effects might be related to the inhibition of sporophytic apomixis in *Citrus*.

- The hypothesis by Ozias-Akins *et al.* (1998) is that ASGR represent a complex locus where gene silencing might be involved as part of the apomixis mechanism (silencing or repression of genes required for normal sexual reproduction).

- Gametophytic and some sporophytic apomixis genes are linked to lethal gametic factors that make inheritance studies of apomixis even more difficult.

All this knowledge and hypothesis must be considered now as a starting point to be complemented by:

- More parents with larger progenies. Different genotypes may present other alleles and other genomic regions might become unveiled.

- Apomixis has to be considered as a continuous variable although not normally distributed which imply that new statistical methodologies for QTL detection have to be developed to allow more

powerful estimation of gene effects and individual contribution of QTLs detected.

- QTL analysis of candidate genes will be very useful because knowing the metabolic function of genes governing apomixis will allow to check hypothesis on the key steps controlling the apomictic process. Good candidate genes will be those related to the synthesis and signal transduction of phytohormones. On the other hand, the recent advancement of Molecular Biology on the isolation of genes putatively related with apomixis (Luo *et al.*, 1999; Chen *et al.*, 1999; Guerin *et al.*, 2000) will be a logical source of candidate genes.

6. ACKNOWLEDGEMENTS

We are grateful to Drs. L. Navarro and M. Roose for their helpful comments and Dr. J.B. Forner for allowing us to use the C x P progeny. This work was supported by research grant INIA (SC95-034).

7. REFERENCES

- Asker, S.E., and Jerling, L. (1992). *Apomixis in Plants*. CRC Press, Boca Raton.
- Bacchi, O. (1943). Cytological observations in Citrus. III. Megasporogenesis, fertilization and polyembryony. *Bot. Gaz.* 105: 221-225.
- Bashaw, E.C. (1975). Problems and possibilities of apomixis in the improvement of tropical forage grasses, in Doll, E.C. and Mott, G.O. (eds.), *Topical forages in livestock production systems*. Spec. Pub. 24. *American Society of Agronomy*. Madison, WI. p. 23-30.
- Bicknell, R.A., Borst, N.K., and Koltunow, A.M. (2000). Monogenic inheritance of apomixis in two *Hieracium* species with distinct developmental mechanisms. *Heredity* 84: 227-237.
- Cameron, J. W., and Garber, M. J. (1986). Identical-Twin Hybrids of *Citrus* x *Poncirus* from strictly sexual seed parents. *Amer. J. Bot.* 55: 199-205.
- Cameron, J. W., and Soost, R. (1979). Sexual K and Nucellar Embryony in F₁ Hybrids and Advanced Crosses of *Citrus* with *Poncirus*. *J. Amer. Soc. Hort. Sci.* 104: 408-410.
- Carman, J.G. (1997). Asynchronous expression of duplicate genes in angiosperms may cause apomixis, bispory, tetraspory, and polyembryony. *Biol. J. Linn. Soc.* 61: 51-94.
- Clausen, J. (1961). Introgression facilitated by apomixis in polyploid poas. *Euphytica* 10: 87-94.
- Chen, L., Miyazaki C., Kojima, A., Saito, A., and Adachi, T. (1999). Isolation and characterization of a gene expressed during early embryo sac development in apomictic Guinea Grass (*Panicum maximum*), *J. Plant Physiol.* 154: 55-62.
- Esan, E.B. (1973) A detailed study of adventive embryogenesis in the Rutaceae. PhD Dissertation. Univ. California, Riverside.
- Fryxell, P. A. (1957). Mode of reproduction of higher plants. *Botan. Rev.* 23: 132-233.
- Furusato, K. (1954). Studies on polyembryony in citrus. *Annual Rept. Natl. Inst. Genet.* Japan. 4.

- García, R., Asíns, M.J., Forner, J., and Carbonell, E.A. (1999). Genetic analysis of apomixis in *Citrus* and *Poncirus* by molecular markers. *Theor. Appl. Genet.* 99: 511-518.
- Guerin, J., Rossel, J.B., Robert, S., Tsuchiya, T., and Koltunow, A. (2000). A DEFICIENS homologue is down-regulated during apomict initiation in ovules of *Hieracium*. *Planta* 210: 914-920.
- Hanna, W.W., and Bashaw, E.C. (1987). Apomixis: its identification and use in plant breeding. *Crop Science* 27: 1136-1139.
- Iwamasa, M., Ueno, I., and Nishiura, M. (1967). Inheritance of nucellar embryony in *Citrus*. *Bull. Hort. Sta. Japan* ser. B. n° 7.
- Kindiger, B., Sokolov, V., and Khatypova, I.V. (1996a). Evaluation of apomictic reproduction in a set of 39 chromosome maize-*Tripsacum* backcross hybrids. *Crop Science* 36: 1108-1113.
- Kindiger, B., Bai, D., and Sokolov, V. (1996b). Assignment of a gene(s) conferring apomixis in *Tripsacum* to a chromosome arm: cytological and molecular evidence. *Genome* 39: 1133-1141.
- Koltunow, A.M., Bicknell, R.A., and Chaudhury, A.M. (1995). Apomixis: Molecular strategies for the generation of genetically identical seeds without fertilization. *Plant Physiol.* 108: 1343-1352.
- Koltunow, A.M. (1993). Apomixis: embryo sacs and embryos formed without meiosis or fertilization in ovules. *The Plant Cell* 5: 1425-1437.
- Lakshmanan, K.K., and Ambegaokar, K.K. (1984). Polyembryony, in Jhori, B.M. (ed.), *Embryology of Angiosperms*, Springer-Verlag, pp. 445-474.
- Leblanc, O., Grimanelli, D., González-de León, and Savidan, Y. (1995a). Detection of the apomictic mode of reproduction in maize-*Tripsacum* hybrids using maize RFLP markers. *Theor. Appl. Genet.* 90: 1198-1203.
- Leblanc, O., Peel, M.D., Carman, J.G., and Savidan, Y. (1995b). Megasporogenesis and megagametogenesis in several *Tripsacum* species (Poaceae). *Am. J. Bot.* 82: 57-63.
- Luo, M., Bilodeal, P., Koltunow, A., Dennis, E.S., Peacock, W.J., and Chaudhury, A.M. (1999). Genes controlling fertilization-independent seed development in *Arabidopsis thaliana*, *Proc. Natl. Acad. Sci. USA* 96: 296-301.
- Mazzucato, A., Falcinelli, M., and Veronesi, F. (1996). Evolution and adaptedness in a facultatively apomictic grass, *Poa pratensis* L. *Euphytica* 92: 13-19.
- Minessey, F.A. (1953). Effect of rootstock on polyembryony in *Citrus*. *Alexandria J. Agr. Res.* 1: 83-89.
- Nogler, G.A. (1984). Gametophytic apomixis, in Jhori B. M. (ed.), *Embryology of Angiosperms*. Springer-Verlag Berlin, pp. 475-518.
- Ozias-Akins, P., Roche, D., and Hanna, W.W. (1998). Tight clustering and hemizygosity of apomixis-linked in *Pennisetum squamulatum* implies genetic control of apospory by a divergent locus that may have no allelic form in sexual genotypes. *Proc. Natl. Acad. Sci. USA* 95: 5127-5132.
- Parlevliet, J.E., and Cameron, J.W. (1959). Evidence on the inheritance of nucellar embryony in *Citrus*. *Proc. Amer. Soc. Hort. Sci.* 74: 252-260.
- Pessino, S.C., Ortiz, J.P.A., Leblanc, O., Valle, C.B., Evans, C., and Hayward, M.D. (1997). Identification of a maize linkage group related to apomixis in *Brachiaria*. *Theor. Appl. Genet.* 94: 439-444.
- Roche, D., Cong, P., Chen, Z., Hanna, W.W., Gustine, D.L., Sherwood, R.T., and Ozias-Akins, P. (1999). An apospory-specific genomic region is conserved between buffelgrass (*Cenchrus ciliaris* L.) and *Pennisetum squamulatum* Fresen. *The Plant Journal* 19: 203-208.
- Taliaferro, C. M., and Bashaw, E. C. (1966). Inheritance and control of obligate apomixis in breeding buffelgrass. *Crop Sci.* 6: 473-476.

Tisserat, B., and Murashige, T. (1977). Effects of ethephon, ethylene, and 2,4-dichlorophenoxyacetic acid on asexual embryogenesis in vitro. *Plant Physiol.* 60: 437-439.

Tisserat, B., Esan, E.B., and Murashige, T. (1978). Somatic embryogenesis in angiosperms, Horticultural Reviews 1-79.

Vielle Calzada, J.P.V., Crane, C.F., and Stelly, D.M. (1996). Apomixis: the asexual revolution. *Science* 274: 1322-1323.

10

USE OF MOLECULAR MARKERS FOR FRUIT CROP IMPROVEMENT

A.R. McCASKILL¹ and J.J. GIOVANNONI^{1,2*}

¹Boyce Thompson Institute for Plant Research, Cornell University, Ithaca, New York, 14853

²USDA Plant, Soil and Nutrition Laboratory, Cornell University, Ithaca, New York, 14853

Corresponding author email, jjg33@cornell.edu

1. INTRODUCTION

Selectable trait-linked markers have been an integral part of many major plant and animal breeding programs for years. The most common types of markers used generally fall into two categories—morphological and molecular. Both types are the result of differences in genotypes or, in simplest terms, differences in DNA sequences. Morphological markers are identifiable, inherited phenotypes that ideally demonstrate tight linkage with a desired characteristic that is typically not as easily scored as the marker. Molecular markers are also inherited and demonstrate linkage with a desired characteristic; however, they are not based on a visible morphological phenotype. Rather, they are based on variations in DNA sequences, which are usually not observed morphologically (unless they reflect functional and observable allelic differences), yet can be visualized by various molecular techniques.

Great strides have been made in the development and efficiency of molecular markers through the years. The earliest molecular markers used were isozymes. These markers are based on changes in protein electrophoresis resulting from charged amino acid substitutions. Although these were effective for preliminary mapping studies, they are generally extremely limited in number. The development of restriction fragment length polymorphism (RFLP) markers represented a major improvement in marker technology (Helentjaris, 1985). RFLPs are based on differences in restriction sites found in DNA. They are visualized via southern blotting and hybridizing with a specific DNA probe. These markers are relatively abundant and reliable.

The newest variety of markers is PCR based. Included in this category are random amplified polymorphic DNAs (RAPDs; Williams et al, 1990), cleaved amplified polymorphic sequences (CAPS; Konieczny and Ausubel, 1993),

amplified fragment length polymorphisms (AFLPs; Vos et al, 1995) and mini- and microsatellites or simple sequence repeats (SSRs; Tautz, 1989; Weber and May 1989). Since all of these are PCR-based, many markers can be analyzed simultaneously in a very short amount of time; moreover, very little DNA is required for the process. RAPDs utilize random primers that amplify random sequences in the genome. This occasionally results in differential amplification of regions that vary in primer site sequence between the individuals in question resulting in polymorphic amplification products (usually presence/absence) that can be visualized via DNA electrophoresis (see Figure 1). Similarly, CAPS and AFLPs utilize random primers, however, these methods incorporate restriction enzyme digests to increase specificity and reproducibility. SSRs are based upon repetitive DNA sequences comprised of a core sequence that is repeated a variable number of times. This type of marker changes relatively quickly through evolution due to slipping of the DNA polymerase during replication, resulting in a change in the number of core repeat units.

Numerous additional uses for DNA markers have been developed through the years including DNA fingerprinting, pedigree analysis and quantitative trait loci (QTL) analysis. DNA fingerprinting in particular has become an invaluable tool for breeding programs. The technique has been used successfully in many fruit breeding protocols (Nybom, 1994). DNA fingerprinting is used frequently for cultivar identification because of its high level of accuracy. The chance probability that two distinct genotypes exhibit identical DNA fingerprint patterns is less than 10^{-3} using minisatellite DNA hybridization (Nybom, 1994). This method has been used for cultivar identification in papaya (Sharon, 1992), mango (Adato et. al, 1995), grape (Bowers et. al, 1993) and citrus (Fang and Roose, 1997). Fingerprinting has also been promoted as a simple and rapid method of documenting the identity of clonal material used in plant breeding programs, apple included (Mulcahy et. al, 1993). Other uses for this technique relative to fruit crops include genetic characterization of species as demonstrated by Lanham and Brennan (1999) with gooseberry (*Ribes grossularia* subgenus *Grossularia*). It was demonstrated that there is limited genetic diversity among twelve genotypes of the European gooseberry, and thus that breeding programs would greatly benefit from the introduction of different germplasm. Similarly, Antonius and Nybom (1994) determined the amount of genetic diversity in wild raspberry (*Rubus idaeus*) populations via fingerprinting. All samples yielded different fingerprints, which suggested that vegetative propagation might not be as common as originally predicted in wild *R. idaeus* populations. Conversely, Antonius and Nybom found that all samples collected from a population of the apomictic blackberry species *R. nessensis* exhibited identical fingerprints. Kraft et al (1996) used fingerprinting protocols for the evaluation of European blackberry species in the hopes of aiding taxonomic classification. They

suggested that fingerprint analysis could be a valuable tool in taxonomy when used in combination with morphological analysis.

Pedigree analysis also has numerous uses in modern breeding projects. It has been used for paternity determination of apple seedlings derived from open pollination, which results in unknown paternity of seedlings (Nybom and Schaal, 1990). Sneller (1994) employed pedigree analysis to study the pattern of diversity (which is reportedly low) among elite North American soybean lines. Sneller found that efforts to increase genetic diversity in soybeans have had little effect. In addition to these applications, this technique can also be used for monitoring of outcrossing rates, detection of gene introgression in breeding programs and analysis of somatic hybridization (Nybom, 1994).

QTL analysis is the process of identifying the number and relative contribution of genes that influence polygenic traits. This is clearly of great importance since many valuable agricultural traits are regulated by more than one gene. The presence of markers associated with QTLs was first reported by Sax in 1923 who demonstrated that a seed coat color marker in beans was associated with the quantitative seed size trait. More recently, markers have been used extensively to simplify QTL analysis. Grandillo and Tanksley (1996) mapped and analyzed tomato QTLs affecting important agronomic and biological traits based on a cross between an elite processing line (*L. esculentum*) and the wild species *L. pimpinellifolium*. It has also been shown that QTL analysis can be used for crop improvement via identification and introduction of advantageous loci from wild relatives (Tanksley and Nelson, 1996). Similarly, Paran et al (1997) performed a cross between cultivated tomato (*L. esculentum*) and a wild relative (*L. cheesmanii*) to identify QTLs influencing morphological characteristics such as plant height and mass, branching, leaf length, number of nodes and first flower bearing node. Molecular markers played vital roles as tools in the identification and mapping of QTLs in these and additional projects far too numerous to list here.

The process of marker-assisted selection has become a routine step in many plant breeding programs due in large part to the development, reliability, and ease of use of molecular markers in crop plants. Since molecular markers are not environmentally regulated and are detectable at all stages of plant growth (that are capable of yielding sufficient DNA for analysis), they can be much more efficient for selection than traditional morphological markers (Mohan et. al, 1997). This can typically result in decreased overall cost, growth time to trait scoring and labor relative to plant production and maintenance. Gu et al (1995) reported that their large scale, PCR-based method of marker-assisted selection required as little as four hours at a cost of less than \$0.40 per sample. Similar

assay time and cost estimates have been reported by numerous groups and especially for PCR-based assay methods.

Edwards and Page (1994) evaluated the benefits of marker-assisted selection vs. phenotypic recurrent selection through the use of a computer simulation. They found that when using markers tightly linked to a small number of defined QTLs, there were very rapid gains in the first two to three years of selection. Hormaza (1999) demonstrated the benefits of marker-assisted selection when combined with *in vitro* embryo culture for cherry breeding. PCR-based marker selection allowed for identification of seedlings with the desired trait while still in culture, eliminating the need to transplant and grow all of the progeny to a significant size before screening and selecting. Garcia et al (1999) were able to identify markers associated with apomixis QTLs in citrus with the aim of assisting selection of future apomictic rootstocks.

Developments in marker-assisted selection have the potential to benefit not only the crop in which the markers are developed, but also in species with similar genomes. Devos et al (1995) showed that gene order in rye and wheat is highly conserved, suggesting that markers would also be conserved and could be used between species. This conservation of marker and gene order in the grasses not only will facilitate the use of molecular tools across related species, but also holds great promise for isolating genes via positional cloning strategies (Mohan et. al, 1997). Colinearity within the Solanaceae family, which includes tomato, potato and pepper, has also been demonstrated (Prince et al, 1993; Tanksley et al, 1992). Comparison of the tomato and potato genomes shows that five major chromosomal inversions occurred during species differentiation (Tanksley et al, 1992). When the tomato map is compared with the genetic map of pepper, 15 obvious chromosome breakage events are identifiable (Prince et al, 1993). However, many of the pepper linkage groups have been shown to correlate to multiple un-linked regions in the tomato genome suggesting numerous additional re-arrangements and translocations within these segments. Within smaller regions however, marker order is conserved (see Figure 2). In summary, comparative genomics of the grasses and Solanaceae suggests that molecular markers can be used across species even though gene order is not conserved across entire chromosomes. The key to efficient cross-species marker utilization is use of tightly-linked markers to the trait of interest from the marker donor species, confirmation of linkage in the marker recipient species, and knowledge of general likelihood of colinearity among the marker donor and recipient species.

Markers also play a vital role in the task of gene mapping and cloning. In several methods of gene isolation, markers are key tools. For example, if colinearity exists between organisms, markers can be used to identify gene homologs

between species. In addition to colinearity, candidate gene queries can be based on conservation of function of gene products or EST database queries. Conservation of function is based on the principle that gene product function can be conserved between very distant organisms. For example, the function of the de-etiolated 2 (DET2) protein, a 5- α -reductase enzyme, is conserved between humans and *Arabidopsis* (Li et. al, 1997). Another example of the conservation of gene function between much more closely related organisms is the ethylene receptors ERS1 and NEVER-RIPE (NR) from *Arabidopsis* and tomato, respectively. These genes function similarly in both species (Wilkinson et al, 1995). A common method of identifying potential homologues in diverse species is through EST database queries. By searching the amino acid sequence of a whole or conserved coding region of gene of interest into sequence databases, cDNAs from other organisms that have significant sequence similarity can be identified.

Several other methods of gene isolation also incorporate the use of molecular markers. For example, the map-based or positional cloning approach employs DNA marker linkage to first localize the gene of interest to a chromosomal region, then markers are used to narrow the region until the gene is isolated. This technique is commonly used to clone genes when the only information known is the mutant phenotype.

2. APPLICATIONS OF DNA MARKERS TO FRUIT RIPENING

Fruit perform two basic functions relative to assisting plants in propagation: 1) to protect developing seeds and 2) to attract organisms that perform seed dispersal. While varying greatly from species to species, ripening can be generalized as the process which results in the change in fruit from a "protective" to "attractant" function (Fray and Grierson, 1993). The ripening process has been defined into two distinct classes—climacteric and non-climacteric (Biale, 1964). Climacteric fruit such as tomato, most stone fruits, apples, and pears, display the "respiratory climacteric" or burst in respiration associated with the onset of the ripening process. This respiratory rise is also associated with a similarly dramatic increase in ethylene biosynthesis in many climacteric species. Non-climacteric fruits such as strawberry and grapes ripen in the absence of increased respiration and generally seem to undergo ethylene-independent ripening. Interestingly, while ethylene is not required, aspects of ripening (such as color development in citrus flavedo) can be enhanced by application of exogenous ethylene to some non-climacteric fruits.

Genetic determinants clearly play an important role in fruiting and ripening. For example, the activation of ripening-related genes by ethylene mediates the ripening process in climacteric fruit. Many genes are involved in ripening, and

mutations have been identified in several of these (Gray et al, 1994). An extensive map containing ripening-related genes has been developed (Figure 3).

Markers and mapping techniques have proven extremely useful for the analysis of ripening and for fruit quality improvement. Fang et al (1997) developed markers linked to a gene controlling fruit acidity in citrus. Molecular markers have also been identified that segregate with a seedless phenotype in grape (Striem et al, 1996). Fabbri et al (1995) developed markers to distinguish between small-fruited olive cultivars, which are grown primarily for oil, and other large-fruited cultivars used for whole and chopped olive products. Hadas et al (1995) used PCR-based markers to show linkage between a gene encoding invertase and sucrose accumulation in tomato fruit from the species *L. hirsutum* and subsequently demonstrated that the invertase sequence in fact is the QTL (Egashira et al, 1999). DNA markers have been developed that can correctly distinguish between apple plants that will produce red-skinned fruit vs. those that yield yellow-skinned fruit (Cheng et al, 1996).

For years tomato has been the model system of choice for studying fruit development and ripening. In 1980, JF Forbes enthusiastically promoted the use of tomato as a model system for molecular biology for the extensive amount of existing genetic knowledge and because of the vast germplasm collection. There are nine species of tomato—*Lycopersicon esculentum* representing the cultivated tomato and *L. pennellii*, *L. cheesmannii*, *L. peruvianum*, *L. chilense*, *L. hirsutum*, *L. parviflorum*, *L. pimpinellifolium* and *L. chmielewskii* representing the wild species. Interbreeding is possible between any of these species, resulting in hybrids that can have a high degree of polymorphism. For these reasons tomato is an ideal model system for traditional and molecular genetic studies.

Many types of recombinant lines have been developed and are frequently used in tomato research. Eshed and Zamir (1995) used introgression lines for fine mapping of yield-associated QTLs. Chetelat et al (1998) used alien addition lines for gene mapping. They also used these lines for investigating the potential for transferring desirable traits from a wild relative.

Numerous morphological and genetic maps have been developed in tomato. The most extensive tomato linkage map based primarily on DNA markers was constructed using a *L. esculentum* x *L. pennellii* cross due to the low genetic variation found in tomato cultivars. The resulting tomato RFLP map contains more than 1000 markers, which are distributed throughout the genome and are separated by an average of 1.2 cM (Tanksley et al, 1992).

The development of genetic maps has aided in the identification and mapping of many ripening mutations (see Figure 3). *Never-ripe* (*Nr*; Yen et al, 1995), *non-ripening* (*nor*; Giovannoni et al, 1995), *high-pigment 1* (*hp-1*; Yen et al, 1997),

high-pigment 2 (*hp-2*; Mustilli et al, 1998) and *Colorless non-ripening* (*Cnr*; Thompson et al, 1999) represent the major ripening mutants that have been placed on the genetic map of tomato. Several of these genes were isolated with the assistance of markers. For example, the *HP-2* gene was isolated via a RFLP marker/candidate gene approach (Mustilli et al, 1998). Both *RIN* and *NOR* were cloned by a map-based strategy that utilized markers and YAC and BAC libraries (Giovannoni et al., unpublished).

Marker-assisted gene isolation techniques are not limited to cloning ripening-related genes from tomato. They have been used extensively for cloning many types of genes, most notably disease resistance genes. *Cladosporium fulvum* (Cf), *Pseudomonas syringae* pv. tomato (Pto) and root-knot nematode (Mi-1) resistance genes have been isolated via marker-based protocols. Both Cf-9 (Jones et al, 1994) and Cf-4 (Takken et al, 1998) were isolated by transposon tagging. Cf-2 (Dixon et al, 1996) was cloned using positional cloning methods. Martin et al (1993) identified the Pto gene using map-based cloning. The Mi-1 gene (Vos et al, 1998) was cloned using an AFLP-based strategy.

Table 1. References for available fruiting species genetic maps and their marker composition

Species	Reference	Markers
Tomato	Tanksley et al, 1992	RFLP, isozyme, morphological
Strawberry	Davis, 1993	RAPD, morphological, isozyme
Blueberry	Qu and Hancock, 1997	RAPD
	Rowland and Levi, 1994	RAPD
Grape	Lodhi et al, 1995	RAPD, RFLP, isozyme
Apple	Conner et al, 1997	RAPD, isozyme, morphological
Peach	Rajapakse et al, 1994	RFLP, RAPD, morphological
	Dirlewanger et al, 1998	morphological, RFLP, RAPD, IMA, AFLP, isozyme
Citrus	Durham et al, 1992	isozyme, RFLP
	Jarrell et al, 1992	isozyme, RFLP
Melon	Wang et al, 1997	AFLP, RAPD, microsatellite
	Baudracco-Arnas and Pitrat, 1996	RFLP, RAPD, isozyme, morphological
Oil palm	Mayes et al, 1997	RFLP
Papaya	Sondur et al, 1996	RAPD
Pepper	Lefebvre et al, 1995	RFLP, RAPD
Sour cherry	Wang et al, 1998	RFLP
Avocado	Sharon et al, 1997	SSR, FAPD, DFP

Molecular maps have been developed for numerous fruiting species—many of which are highly heterozygous (R. Socias, 1998) allowing for ease of mapping

with molecular markers. Maps available for several species are listed in Table 1. Information gathered about one species may be useful for manipulating fruit quality in another species depending on conservation of gene sequence and order as mentioned above.

3. CONCLUSION

Molecular markers have become invaluable tools for plant breeders in recent years. In breeding programs, they eliminate the need to grow plants to adulthood to select for the desired trait. The result is a dramatic reduction in time, money and space needed—this is particularly important for woody plants that have a long juvenile period. Molecular markers have also become indispensable tools for rapidly and reliably distinguishing between cultivars. In addition, markers have greatly improved the efficiency of mapping and cloning of ripening-related as well as other types of genes. Genetic maps have the potential to aid in identifying and isolating homologous genes in other species as a result of colinearity. As more and more ripening-related genes are identified and characterized, a clearer picture of the phenomenon of fruit ripening and regulation of gene expression will develop, as will a greater knowledge of how ripening regulation and processes have evolved among plant species.

4. REFERENCES

- Adato A, Sharon D, Lavi U, Hillel J, Gazit S (1995) Application of DNA fingerprinting for identification and genetic analyses of mango (*Mangifera indica*) genotypes. *J Am Soc Hortic Sci* 120:259-264
- Antonius K, Nybom H (1994) DNA fingerprinting reveals significant amounts of genetic variation in a wild raspberry *Rubus idaeus* population. *Mol Eco* 3:177-180
- Baudracco-Arnas S, Pitrat M (1996) A genetic map of melon (*Cucumis melo* L.) with RFLP, RAPD, isozyme, disease resistance and morphological markers. *Theor Appl Genet* 93:57-64
- Biale JB (1964) Growth, maturation and senescence in fruits. *Science* 146:880-888
- Bowers J, Bandman E, Meredith C (1993) DNA fingerprint characterization of some wine grape cultivars. *Am J. Enol Vitic* 44:266-284
- Cheng FS, Weeden NF, Brown SK (1996) Identification of co-dominant RAPD markers tightly linked to fruit skin color in apple. *Theor Appl Genet* 93:222-227
- Chetelat RT, Rick CM, Cisneros P, Alpert KB, DeVerna JW (1998) Identification, transmission and cytological behavior of *Solanum lycopersicoides* Dun. Monosomic alien addition lines in tomato (*Lycopersicon esculentum* Mill.). *Genome* 41:40-50
- Conner PJ, Brown SK, Weeden NF (1997) Randomly amplified DNA-based genetic linkage maps of three apple cultivars. *J Am Soc Hortic Sci* 122:350-359
- Davis TM, Yu H (1997) A linkage map of the diploid strawberry, *Fragaria vesca*. *J Heredity* 88:215-221
- Devos K, Moore G, Gale M (1995) Conservation of marker synteny during evolution. *Euphytica* 85:367-372
- Dirlewanger E, Pronier V, Parvery C, Rothan C, Guye A, Monet R (1998) Genetic linkage map of peach [*Prunus persica* (L.) Batsch] using morphological and molecular markers. *Theor Appl Genet* 97:888-895

- Dixon M, Jones D, Keddle J, Thomas C, Harrison K, Jones J (1996) The tomato Cf-2 disease resistance locus comprises two functional genes encoding leucine-rich repeat proteins. *Cell* 84:451-459
- Doganlar S, Tanksley S, Mutschler M (2000) Identification and molecular mapping of loci controlling fruit ripening time in tomato. *Theor Appl Genet* 100:249-255
- Dudley JW (1993) Molecular markers in plant improvement: manipulation of genes affecting quantitative traits. *Crop Sci* 33:660-668
- Durham RE, Liou PC, Gmitter FG JR, Moore GA (1992) Linkage of restriction fragment length polymorphisms and isozymes in citrus. *Theor Appl Genet* 84: 39-48
- Edwards MD, Page NJ (1994) Evaluation of marker-assisted selection through computer simulation. *Theor Appl Genet* 88:376-382
- Egashira H, Takahashi S, Doi H, Nishizawa T, Escalante A, Takashina T, Imanishi S (1999) Genetic analysis of sucrose-accumulating ability in *Lycopersicon peruvianum*. *Breed Sci* 49:155-159
- Eshed Y, Zamir D (1995) An introgression line population of *Lycopersicon pennellii* in the cultivated tomato enables the identification and fine mapping of yield-associated QTL. *Genetics* 141:1147-1162
- Fabbri A, Hormaza JJ, Polito VS (1995) Random amplified polymorphic DNA analysis of olive (*Olea europaea* L.) cultivars. *J Am Soc Hortic Sci* 120:538-542
- Fang DQ, Federici CT, Roose ML (1997) Development of molecular markers linked to a gene controlling acidity in citrus. *Genome* 40:841-849
- Fang DQ, Roose ML (1997) Identification of closely related citrus cultivars with inter-simple sequence repeat markers. *Theor Appl Genet* 95:408-417
- Fray R, Grierson D (1993) Molecular genetics of tomato fruit ripening. *Trends Genet* 9:438-443
- Garcia R, Asins MJ, Forner J, Carbonell EA (1999) Genetic analysis of apomixis in *Citrus* and *Poncirus* by molecular markers. *Theor Appl Genet* 99:511-518
- Giovannoni JJ (1993) Molecular and genetic analysis of tomato fruit development and ripening, in J. Bryant (ed.), *Methods in Plant Biochemistry vol. 10*, Academic Press Limited, London, pp. 251-285
- Giovannoni JJ, Noensie E, Ruezinksy D, Lu X, Tracy S, Ganai M, Martin G, Pillen K, Alpert K, Tanksley S (1995) Molecular genetic analysis of the *ripening-inhibitor* and *non-ripening* loci of tomato: a first step in genetic map-based cloning of fruit ripening genes. *Mol Gen Genet* 248:195-206
- Giovannoni J, Yen H, Shelton B, Miller S, Vrebalov J, Kannan P, Tieman D, Hackett R, Grierson D, Klee H (1999) Genetic mapping of ripening and ethylene-related loci in tomato. *Theor Appl Genet* 98:1005-1013
- Goodfellow PN, Sefton L, Farr CJ (1993) Genetic maps *Phil Trans R Soc Lond B* 339:139-146
- Grandillo S, Ku H-M, Tanksley SD (1996) Characterization of *fs8.1*, a major QTL influencing fruit shape in tomato. *Mol Breed* 2:251-260
- Grandillo S, Tanksley SD (1996) QTL analysis of horticultural traits differentiating the cultivated tomato from the closely related species *Lycopersicon pimpinellifolium*. *Theor Appl Genet* 92:935-951
- Gray JE, Picton S, Giovannoni JJ, Grierson D (1994) The use of transgenic and naturally occurring mutants to understand and manipulate tomato fruit ripening. *Plant Cell Environ* 17:557-571
- Gu WK, Weeden NF, Yu J, Wallace DH (1995) Large-scale, cost-effective screening of PCR products in marker-assisted selection applications. *Theor Appl Genet* 91:465-470
- Hadas R, Schaffer A, Miron D, Fogelman M, Granot D (1995) PCR-generated molecular markers for the invertase gene and sucrose accumulation in tomato. *Theor Appl Genet* 90:1142-1148
- Helentjaris T, King G, Slocum M, Siedenstrang C (1985) Restriction fragment polymorphisms as probes for plant diversity and their development as tools for applied plant breeding. *Plant Mol Biol* 5: 109-118
- Hormaza JI (1999) Early selection in cherry combining RAPDs with embryo culture. *Sci Hortic* 79:121-126

- Jarrell DC, Roose ML, Traugh SN, Kupper RS (1992) A genetic map of citrus based on the segregation of isozymes and RFLPs in an intergeneric cross. *Theor Appl Genet* 84:49-56
- Kijas JMH, Thomas MR, Fowler JCS, Roose ML (1997) Integration of trinucleotide microsatellites into a linkage map of *Citrus*. *Theor Appl Genet* 94:701-706
- Konieczny A, Ausubel FM (1993) A procedure for mapping *Arabidopsis* mutations using co-dominant ecotype specific PCR-based markers. *Plant J* 4:403-410
- Kraft T, Nybom H, Werlemark G (1996) DNA fingerprint variation in some blackberry species (*Rubus* subg. *Rubus*, *Rosaceae*). *Pl Syst Evol* 199:93-108
- Lanham PG, Brennan RM (1999) Genetic characterization of gooseberry (*Ribes grossularia* subgenus *Grossularia*) germplasm using RAPD, ISSR and AFLP markers. *J Hortic Sci Biotech* 74:361-366
- Law CN, Worland AJ (1996) Inter-varietal chromosome substitution lines in wheat—revisited. *Euphytica* 89:1-10
- Lefebvre V, Palloix A, Caranta C, Pochard E (1995) Construction of an intraspecific integrated linkage map of pepper using molecular markers and doubled-haploid progenies. *Genome* 38:112-121
- Lelievre J-M, Latche A, Jones B, Bouzayen M, Pech J-C (1997) Ethylene and fruit ripening. *Physiol Plant* 101:727-739
- Lodhi, MA, Daly MJ, Ye G-N, Weeden NF, Reisch BI (1995) A molecular marker based linkage map of *Vitis*. *Genome* 38:786-794
- Manning K (1998) Isolation of a set of ripening-related genes from strawberry: their identification and possible relationship to fruit quality traits. *Planta* 205:622-631
- Martin G, Brommonschenkel S, Chunwongse J, Frary A, Ganai M, Spivey R, Wu T, Earle E, Tanksley S (1993) Map-based cloning of a protein kinase gene conferring disease resistance in tomato. *Science* 262:1432-1436
- Mayes S, Jack PL, Marshall DF, Corley RHV (1997) Construction of a RFLP genetic linkage map for oil palm (*Elaeis guineensis* Jacq.). *Genome* 40:116-122
- Mohan M, Nair S, Bhagwat A, Krishna TG, Yano M, Bhatia CR, Sasaki T (1997) Genome mapping, molecular markers and marker-assisted selection in crop plants. *Mol Breed* 3:87-103
- Mulcahy DL, Cresti M, Sansavini S, Douglas GC, Linskens HF, Mulcahy GB, Vignani R, Pancaldi M (1993) The use of random amplified polymorphic DNAs to fingerprint apple genotypes. *Sci Hortic* 54:89-96
- Mustilli A-C, Fenzi F, Ciliento R, Alfano F, Bowler C (1999) Phenotype of the tomato *high pigment-2* mutant is caused by a mutation in the tomato homolog of DEETIOLATED 1. *Plant Cell* 11:145-157
- Nybom H (1994) DNA fingerprinting—a useful tool in fruit breeding. *Euphytica* 77:59-64
- Nybom H, Schaal BA (1990) DNA fingerprints applied to paternity analysis in apples *Malus domestica*. *Theor Appl Genet* 79:763-768
- Paran I, Goldman I, Tanksley SD, Zamir D (1995) Recombinant inbred lines for genetic mapping in tomato. *Theor Appl Genet* 90:542-548
- Paran I, Goldman I, Zamir D (1997) QTL analysis of morphological traits in a tomato recombinant inbred line population. *Genome* 40:242-248
- Prince JP, Pochard E, Tanksley SD (1993) Construction of a molecular linkage map of pepper and a comparison of synteny with tomato. *Genome* 36:404-417
- Qu L, Hankcock JF (1997) Randomly amplified polymorphic DNA- (RAPD-) based genetic linkage map of blueberry derived from an interspecific cross between diploid *Vaccinium darrowi* and tetraploid *V. corymbosum*. *J Am Soc Hortic Sci* 122:69-73
- Rajapakse S, Belthoff LE, He G, Estager AE, Scorza R, Verde I, Ballard RE, Baird WV, Callahan A, Monet R, Abbott AG (1995) Genetic linkage mapping in peach using morphological, RFLP and RAPD markers. *Theor Appl Genet* 90:503-510
- Rowland LJ, Levi A (1994) RAPD-based genetic linkage map of blueberry derived from a cross between diploid species (*Vaccinium darrowi* and *V. elliotii*). *Theor Appl Genet* 87:863-868
- Sax K (1923) The association of size differences with seed coat pattern and pigmentation in *Phaseolus vulgaris*. *Genetics* 8:552-560

- Sharon D, Cregan PB, Mhameed S, Kusharska M, Hillel J, Lahav E, Lavi U (1997) An integrated genetic linkage map of avocado. *Theor Appl Genet* 95:911-921
- Sharon D, Hillel J, Mhameed S, Cregan P, Lahav E, Lavi U (1998) Association between DNA markers and loci controlling avocado traits. *J Am Soc Hort Sci* 123:1016-1022
- Sharon D, Hillel J, Vainstein A, Lavi U (1992) Application of DNA fingerprints for identification and genetic analysis of *Carica papaya* and other *Carica* species. *Euphytica* 62:119-126
- Sneller CH (1994) Pedigree analysis of elite soybean lines. *Crop Sci* 34:1515-1522
- Socias, R (1998) Fruit tree genetics at a turning point: the almond example. *Theor Appl Genet* 96:588-601
- Sondur SN, Manshardt RM, Stiles JI (1996) A genetic linkage map of papaya based on randomly amplified polymorphic DNA markers. *Theor Appl Genet* 93:547-553
- Striem MJ, Ben-Hayyim G, Spiegel-Roy, P (1996) Identifying molecular genetic markers associated with seedlessness in grape. *J Amer Soc Hort Sci* 121:758-763
- Takken F, Schipper D, Nijkamp HJJ, Hillel J (1998) Identification and Ds-tagged isolation of a new gene at the *Cf-4* locus of tomato involved in disease resistance to *Cladosporium fulvum* race 5. *Plant J* 14:401-411
- Tanksley SD, Ganai MW, Prince JP, de Vicente MC, Bonierbale MW, Broun P, Fulton TM, Giovannoni JJ, Grandillo S, Martin GB, Messeguer R, Miller JC, Miller L, Paterson AH, Pineda O, Roder MS, Wing RA, Wu W, Young ND (1992) High density molecular linkage maps of the tomato and potato genomes. *Genetics* 132:1141-1160
- Tanksley SD, Nelson JC (1996) Advanced backcross QTL analysis: a method for the simultaneous discovery and transfer of valuable QTLs from unadapted germplasm into elite breeding lines. *Theor Appl Genet* 92:191-203
- Tautz D (1989) Hypervariability of simple sequences as general source for polymorphic DNA markers. *Nucleic Acids Res* 17:6463-6472
- Thompson A, Tor M, Barry C, Vrebalov J, Orfila C, Jarvis M, Giovannoni J, Grierson D, Seymour G (1999) Molecular and genetic characterization of a novel pleiotropic tomato-ripening mutant. *Plant Physiol* 120:383-389
- Vos P, Hogers R, Bleeker M, Reijans M, van de Lee T, Hornes M, Frifiters A, Pot J, Peleman J, Kuiper M, Xabeau M (1995) AFLP: A new technique for DNA fingerprinting. *Nucleic Acids Res* 23:4407-4414
- Vos P, Simons G, Jesse T, Wijbranke J, Heinen L, Hogers R, Frijters A, Groenenkijk J, Diergaarde P, Reijans M, Fierens-onstenk J, de Both M, Peleman J, Liharska T, Hontelez J, Zabeau M (1998) The tomato *Mi-1* gene confers resistance to both root-knot nematodes and potato aphids. *Nature Biotech* 16:1365-1369
- Wang D, Karle R, Brettin TS, Iezzoni AF (1998) Genetic linkage map in sour cherry using RFLP markers. *Theor Appl Genet* 97:1217-1224
- Wang Y-H, Thomas CE, Dean RA (1997) A genetic map of melon (*Cucumis melo* L.) based on amplified fragment length polymorphism (AFLP) markers. *Theor Appl Genet* 95:791-798
- Weber JJ, May PE (1989) Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. *Am J Hum Genet* 44:388-396
- Weeden NF, Hemmat M, Lawson DM, Lodhi M, Bell RL, Manganaris AG, Reisch BI, Brown SK, Ye G-N (1994) Development and application of molecular marker linkage maps in woody fruit crops. *Euphytica* 77:71-75
- Wilkinson J, Lanahan M, Yen HC, Giovannoni J, Klee H (1995) An ethylene-inducible component of signal transduction encoded by Never-ripe. *Science* 270:1807-1809
- Williams JGK, Kubelik AR, Livak KJ, Rafalski JA, Tingey SV (1990) DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. *Nucleic Acids Res* 18: 6531-6536
- Yen HC, Lee S, Tanksley S, Lanahan M, Klee H, Giovannoni J (1995) The tomato *Never-ripe* locus regulates ethylene-inducible gene expression and is linked to a homolog of the *Arabidopsis ETR1* gene. *Plant Physiol* 107:1343-1353

- Yen HC, Shelton BA, Howard LR, Lee S, Vrebalov J, Giovannoni JJ (1997) The tomato *high-pigment* (*hp*) locus maps to chromosome 2 and influences plastome copy number and fruit quality. *Theor Appl Genet* 95:1069-1079
- Yen Y, Baenziger PS (1992) A better way to construct recombinant chromosome lines and their controls. *Genome* 35:827-830
- Zegzouti H, Jones B, Frasse P, Marty C, Maitre B, Latche A, Pech J-C, Bouzayen M (1999) Ethylene-regulated gene expression in tomato fruit: characterization of novel ethylene-responsive and ripening-related genes isolated by differential display. *Plant J* 18:589-600

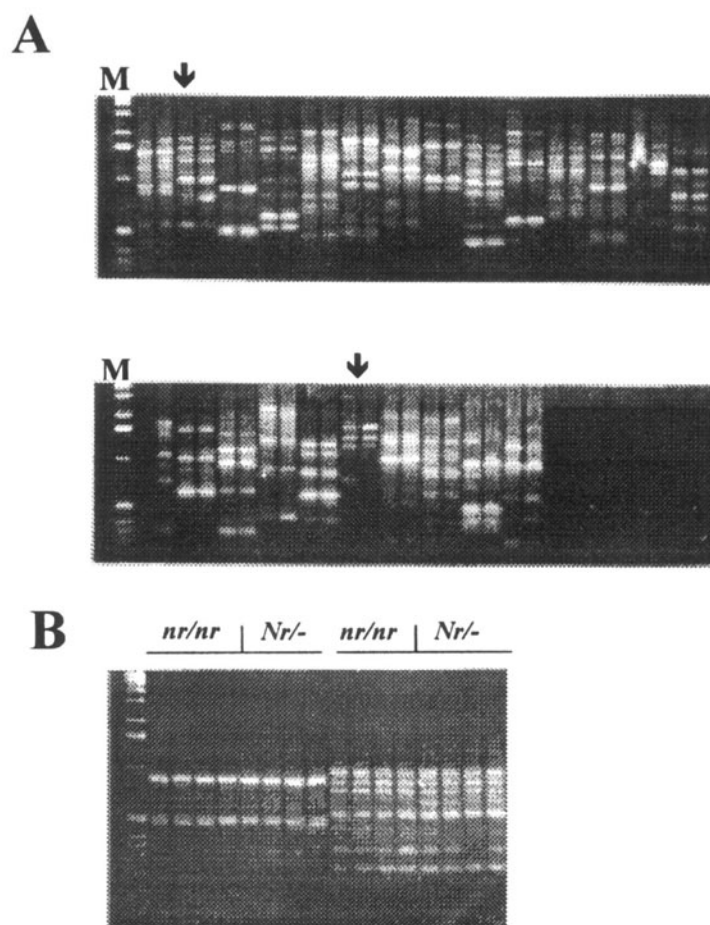


Figure 1. Example of RAPD Marker Identification and Linkage. An F2 population of tomato resulting from the primary cross *L. esculentum* (*Nr/Nr*) X *L. cheesmannii* (*nr/nr*) was grown and scored for seedling ethylene sensitivity as described in Yen et al. (1995). Four normal (*nr/nr*) and four mutant (*Nr/-*) individuals were selected and isolated for genomic DNA. Genomic DNAs were pooled based on phenotype and resulting pooled DNAs were amplified with 10 base random primers. **A** shows 24 pairs of normal (left) and mutant (right) pooled DNA RAPD PCR amplification products using 24 distinct random primers, respectively. The lanes labeled "M" contain the 1 kb DNA marker from Bio-Rad and two pool pairs showing putative RAPD polymorphisms are indicated with arrows. The first lane in the lower panel of **A** contains a failed PCR reaction. **B** shows PCR results from two additional RAPD primers (not in **A**) using the individual F2 genomic DNAs used to create the pooled DNAs in **A**. The primer used in the left experiment shows no linkage to the *Nr* locus as no amplification product is unique to the members of either pool, while the primer used in the right experiment results in an amplification product derived from a locus clearly linked to *Nr*. In summary, a RAPD screen as in **A** is followed by confirmation of putative linked amplification products via testing in a small population of individuals as in **B**.

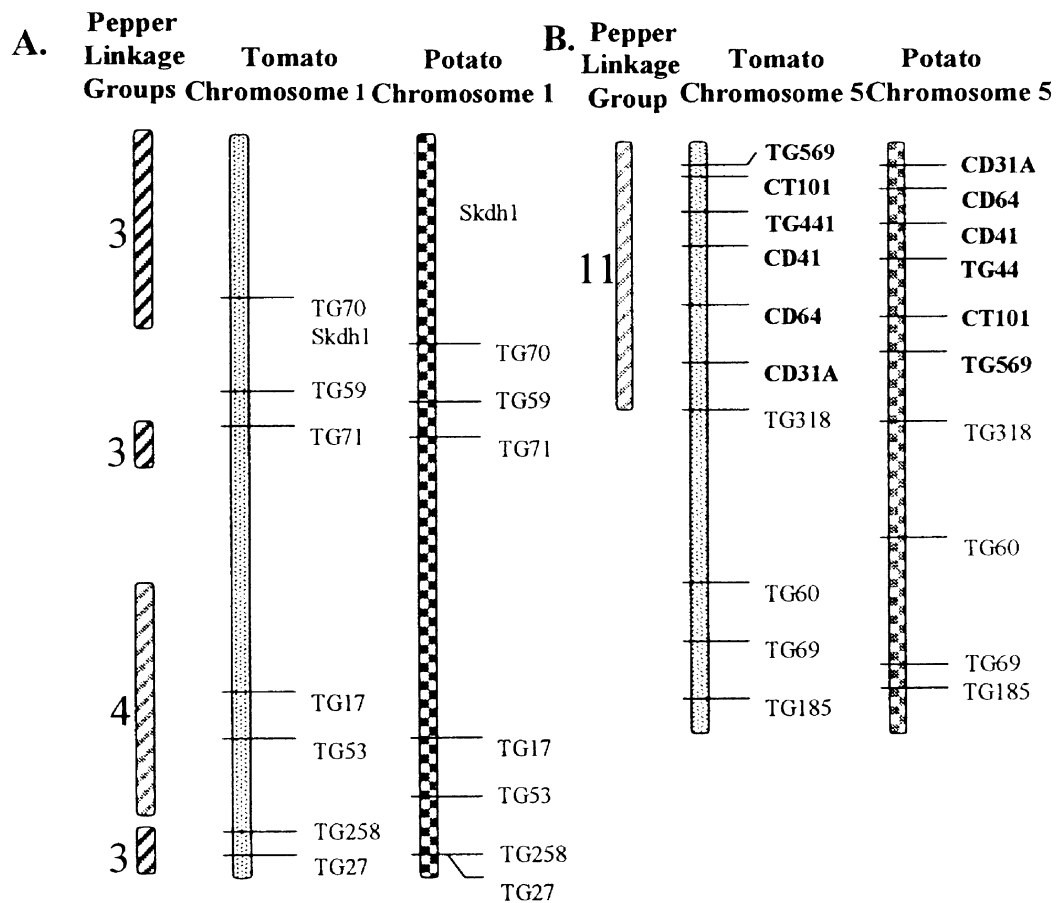


Figure 2. Comparisons of marker order between tomato, potato and pepper. **A.** Marker order is clearly conserved between tomato and potato on Chromosome 1. The hatched bars to the left represent the pepper linkage groups that correspond to areas of the tomato map (Prince et al, 1993). **B.** Illustration of a chromosomal inversion event that occurred between tomato and potato chromosome 5. Markers in bold show where the inverted segment is located. The hatched segment to the left once again represents a pepper linkage group that corresponds to that area of the tomato genome.

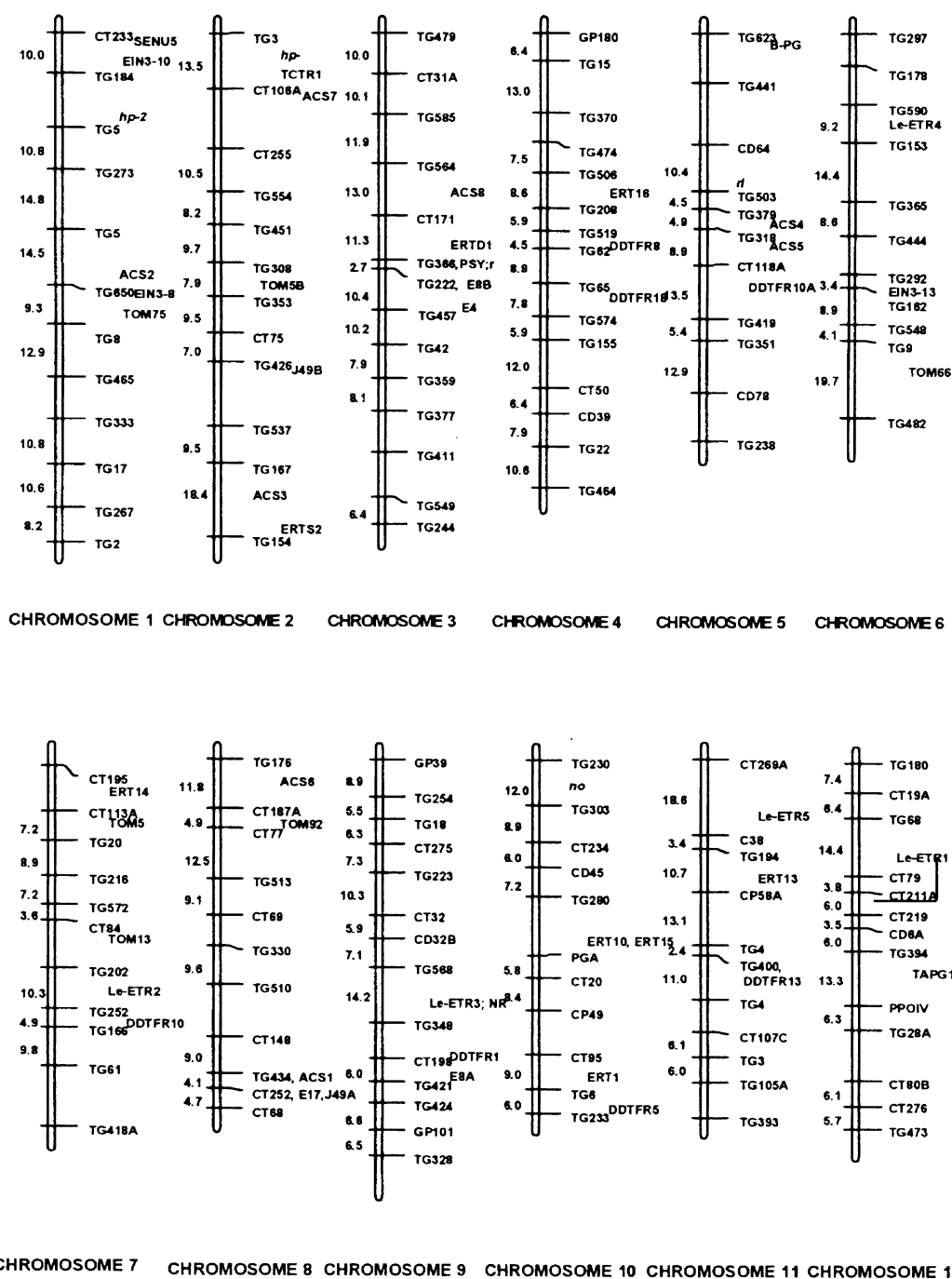


Figure 3. Position of mapped fruit-ripening and ethylene-related loci on the tomato RFLP map (Giovannoni et al., 1999)

11.

***IN SITU* HYBRIDIZATION IN PLANTS – METHODS AND APPLICATION**

Jolanta Maluszynska

*Department of Plant Anatomy and Cytology, University of Silesia,
Jagiellonska 28, 40-032 Katowice, Poland e-mail: maluszyn@us.edu.pl*

1. INTRODUCTION

Cytogenetics is complementary to genetic and molecular analysis of plant genome structure and function. Since its beginning it has been mainly used for identification of chromosomes in plant genomes and karyotype construction (karyotyping). The cytological techniques applied to karyotyping should be easy, reproducible, and should be able to identify major structural rearrangements that can occur between chromosomes. Most significant for the progress in plant cytogenetics was the development of the squash method for chromosome preparation (Darlington, 1937), chromosome banding techniques (Caspersson *et al.*, 1968) and *in situ* hybridization (ISH) techniques (Pardue and Gall, 1969), especially fluorescent *in situ* hybridization (FISH) (Pinkel *et al.*, 1986). These methods have had a major impact on investigation and understanding of plant genome organisation.

Karyotype analysis is usually based on mitotic metaphase chromosomes. The chromosome number (2n), known for angiosperms, range from 4 to 600 (Bennett, 1998). The chromosomes can differ in size and morphology between and within species, and can be identified on the basis of their physical characteristics (length, position of centromere, arm ratio, secondary construction). Very often chromosomes are numerous, small and poorly differentiated in morphology. Their identification needs utilization of additional markers specific for particular chromosomes. Application of differential chromosome staining techniques reveals regions of heterochromatin, which can be recognised as bands or dots. Chromosome banding techniques, such as C-banding and N-banding, were successfully used for cereal karyotyping, identification of chromosome substitution, elimination or translocation and for studying interspecific relationships (Lukaszewski and Gustafson, 1983; Gustafson, 1983; Linde-Laursen and Bothmer, 1988; for review see Lange, 1988). However, they have had limited application in

cytogenetics of plants with small chromosomes. Therefore decreased interest in plant cytogenetics and domination of molecular genetic approaches have been observed over the last two decades.

A significant interest in plant cytogenetics has occurred in recent years with the introduction of molecular cytogenetic techniques for the analysis of somatic and meiotic chromosomes. The main technique, ISH with its different modifications such as FISH, GISH (genomic *in situ* hybridization), mFISH (multicolour fluorescent *in situ* hybridization) is used for detection of chromosomal DNA in cytological preparations. These methods allow for physical gene mapping and localization of different non-coding DNA sequences in chromosomes or interphase nuclei. Also, gene expression can be investigated with the RNA-RNA *in situ* hybridization by localization of RNA as a product of active genes. Specific proteins involved in structure and function of chromosomes can be detected with immunostaining methods.

This review will present basic methods of molecular cytogenetics and their current implications for fundamental or practical investigations. Application of these methods will be focused on DNA-DNA ISH in crop plants with large and small genomes, as well as in model plant species such as *Arabidopsis thaliana* L.

2. *IN SITU* HYBRIDIZATION METHOD

The methods of *in situ* hybridization link molecular biology with cytology by using tools of molecular biology to a microscope preparation of chromosomes, nuclei, cells or tissue. It is a bridge between the molecular and cellular level of genome investigations. This method bases on annealing a single-stranded, labelled DNA or RNA probe with a single-strand nucleic acid in the cell. DNA-DNA ISH allows localization of DNA sequences in chromosomes and interphase nuclei. RNA ISH shows gene expression patterns in cells or tissues.

Early *in situ* studies used radioactively labelled probes and autoradiographic detection of hybridization sites (Pardue and Gall, 1969). Since 1982 several nonisotopic techniques have been developed using chemically modified DNA probes and cytochemical or fluorescent detection (Langer-Safer *et al.*, 1982). The most effective for plant chromosomes are systems that use biotin (Rayburn and Gill, 1985) or digoxigenin (Leitch *et al.*, 1991a, b) labelled nucleotides detected by antibodies conjugated with enzymes (e.g. horseradish peroxidase; alkaline phosphatase) or fluorophores (e.g. fluorescein, rhodamine, Cy3). This ISH technique has been simplified with the use of direct fluorescent DNA labelling, which does not require any detection procedure. Currently, DNA *in situ* hybridization has become almost synonymous with FISH.

FISH has several advantages over previous ISH technologies: (1) does not require a special laboratory for work with radioactivity, (2) decrease the

time required for detection, (3) improve resolution, (4) decrease nonspecific background, and (5) provides the possibility for simultaneous use of different fluorescence systems. Improvements in technology over the last years, such as the epifluorescence microscope with multiple wavelength band-pass fluorescence filters, computer image analysis, and automated controllers of denaturation and hybridization time and temperature, based on thermal cycling technology, increased detection sensitivity have made the FISH technique a routine tool in plant genome analysis. Many detailed protocols of the use of this technique for various plant species have been published and presented in special manuals e.g.: “Plant Chromosomes: Laboratory Methods” (Fukui and Nakayama, 1996); “Plant chromosomes” (Sharma and Sharma, 1999); “Practical *in situ* Hybridization” (Schwarzacher and Heslop-Harrison, 2000). The principles of the method will be briefly presented (Fig. 1).

2.1. Procedure

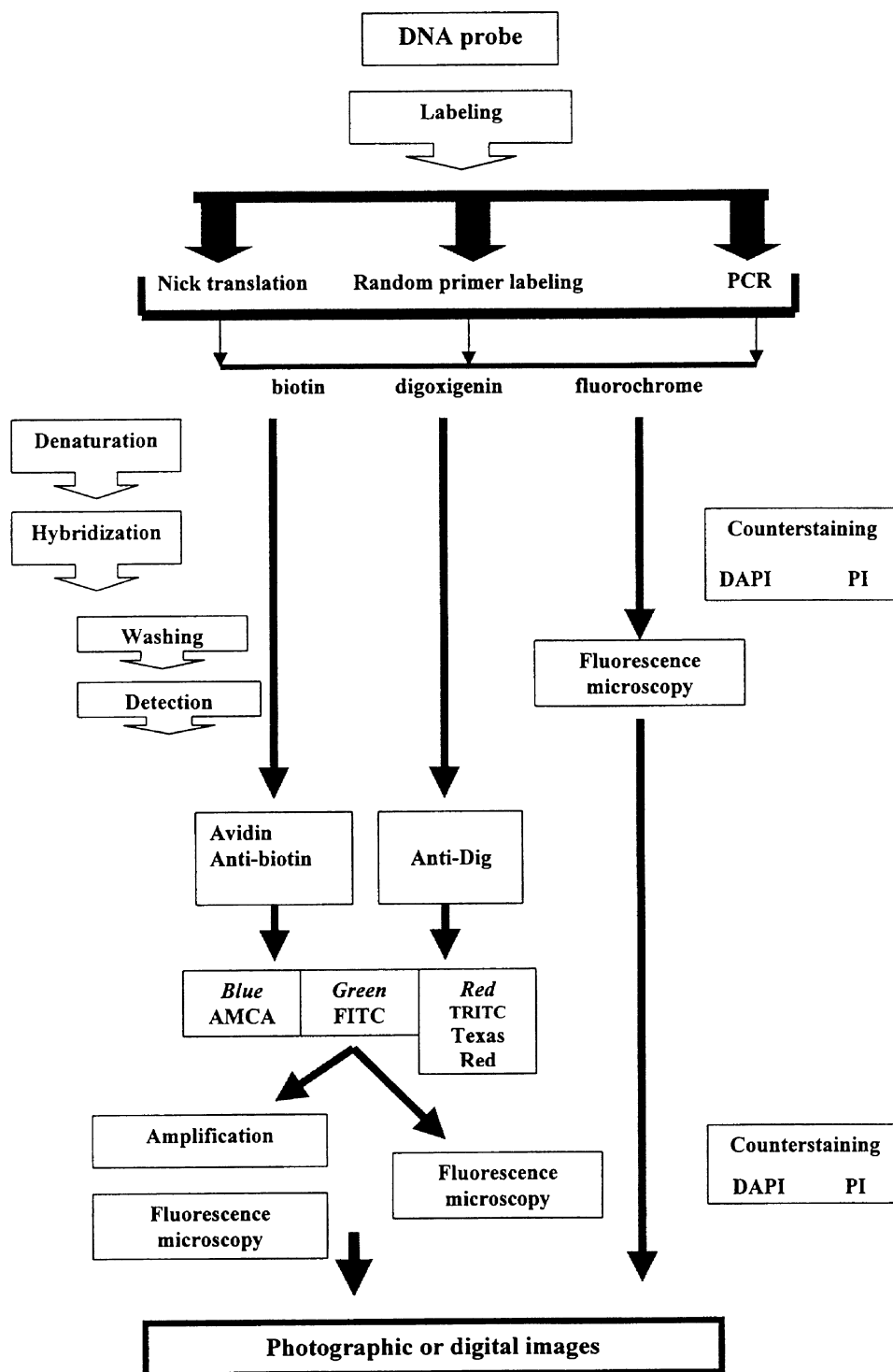
2.1.1. DNA probe

The probe can be cloned DNA, synthetic oligonucleotide or total genomic DNA labelled with haptens like biotin, digoxigenin or fluorophores. Labelled nucleotide analogs (e.g.: biotin-11-dUTP, digoxigenin-11-dUTP, fluorescein-11-dUTP) can be incorporated by nick translation, random primer labelling, end tailing or by polymerase chain reaction (PCR), according to supplier procedure. The DNA used as a probe can be repetitive, as well as low-copy DNA sequences. Tandemly repeated and dispersed sequences have been the targets detected by ISH. Coding, noncoding DNA sequences and mobile elements as well have been used. Sensitivity of the detection depends on the size of targets. The most convenient for detection are tandemly repeated relatively long sequences like the rRNA gene family. However, targets as small as 1-3 kb have been detected in plant genome (Fransz *et al.*, 1996b; Moscone *et al.*, 1996; Ohmido *et al.*, 1998).

2.1.2. Target material

In situ hybridization can be applied to chromosome spreads or mounted and sectioned tissue. Most mitotic plant chromosome studies have been done on root tip preparations after enzymatic digestion of the cell wall. Pollen mother cells have been used for analysis of meiotic chromosomes. High quality chromosome preparations are necessary for good result of *in situ* hybridization. Well-spread, very flat and cytoplasm-free preparations give good chromosome morphology and the highest hybridization signals.

DNA, and particularly RNA ISH are most often performed on sections adhered to microscope slides that are cut from paraffin wax or resin embedded material or prepared after freeze drying. Depending on the required



1. Figure Scheme of fluorescent DNA *in situ* hybridization

resolution, the sections are cut using a microtome, ultramicrotome or vibratome.

2.1.3. Hybridization

To allow hybridization of the labelled DNA probe with the target, they both must be denaturated to make them single-stranded. Conditions of denaturation have to be established empirically by adjusting time, temperature and/or concentration of formamide and salt solutions. For hybridization, a single-stranded probe (denaturation 5 min in 90°C) is applied to denaturated target on the slide (e.g. 5 min at 70-80°C) and incubated to allow the probe to find homologous sequences on the chromosomes or nuclei. Most protocols involve an overnight (about 16 h at 37°C) hybridizations but sometime it is prolonged to a few days. The forming of hybrid (double helix) molecules depends on the level of homology between the probe and target. Hybrids being complementary remain stable, but all weak probe-target complexes and unhybridized probes have to be washed away (stringent washes). After post-hybridization washes the hybridization sites have to be detected and visualized.

2.1.4. Detection

Depending on the probe labelling, various options are available for detection of hybridization sites. For the DNA-DNA hybridization, a fluorescent detection is most frequently used. Occasionally enzyme-linked detection reagents are applied. This is an especially useful method when an epifluorescent microscope is not available (Weiss *et al.*, 1999). The most convenient is the use of probes directly labelled with fluorophores. Slides, directly after post hybridization washes, can be analysed with a microscope. Labelling with biotin or digoxigenin requires detection with avidin or antibodies (antibiotin, antidigoxigenin) which are conjugated with different fluorophores. This method has an additional advantage as hybridization signals can be amplified with primary and secondary antibodies linked with fluorophore. Amplification of signals is specially important when single- or low-copy number sequences are detected.

2. 1. 5. Analysis of signals and imaging

Results of ISH are analysed with a light microscope, when non-fluorescence detection is applied or with an epifluorescent microscope with the appropriate filters for fluorescent detection system. Signals of ISH can be recorded on film or electronically with a CCD camera. The most effective way of data visualisation and presentation is image processing with suitable software, e.g. Adobe Photoshop.

2. 2. Types

Depending on modifications of the ISH procedure the following types can be distinguished:

FISH - fluorescent *in situ* hybridization – with the use of fluorescent labels or detection reagents

mFISH - multicolour FISH - when different fluorescence systems are simultaneously used for more than one probe

PRINS – primed *in situ* hybridization –when thermostable polymerase is used to extend the DNA from sites of oligonucleotide hybridization

Pachytene-FISH – when FISH is applied to meiotic, pachytene chromosomes

EDF–FISH – extended DNA fibres FISH – when detection of DNA sequences is on extended DNA fibres

GISH – genomic *in situ* hybridization – when total genomic DNA is used as a probe.

All these techniques are extremely valuable for the investigation of plant genome structure and function. They enable the localization of different DNA sequences not only on mitotic or meiotic chromosomes, but also in interphase nuclei when genes are transcriptionally active and the DNA is replicated. The development of a large spectrum of various ISH methods stimulates research currently known as comparative cytogenetic genome analysis. This research can reveal phylogenetic relationships between species and chromosomal changes taking place during evolution. These studies have important implications for basic research and practical applications.

3. MARKERS FOR CHROMOSOME IDENTIFICATION

Chromosome number and morphology is characteristic for each species and named karyotype. The base for karyotype construction (karyotyping) is the identification of an individual chromosome or pair of homologous chromosomes. The karyotype can be changed during a plant breeding program involving chromosome engineering or biotechnology procedures like plant transformation or cell and tissue culture *in vitro*. The ability to identify chromosomes and detect any rearrangement is extremely important for the understanding of processes, which occurred during genetic manipulations.

In most plants chromosomes are difficult to identify because of their similar size and lack of morphological markers. The distribution pattern of tandemly repeated DNA sequences can be a good marker for chromosomes or chromosome arms. Tandemly repeated DNA sequences form a major component of plant genomes and usually are species- or chromosome-specific.

Some tandem repeats are coding, like ribosomal RNA genes, while others are non-coding like telomeric DNA and satellite DNA families. They are usually located in heterochromatin region of chromosomes. The majority of genes present in a single or low copy number per genome are usually chromosome-specific. However, single-copy ISH to plant chromosomes is more difficult and is still not routine. Nevertheless, some laboratories have overcome this problem and are successful in localization of particular genes by ISH (see Gene mapping). Retrotransposons, which are another significant component of the genome, can also be suitable chromosome markers. Numerous retrotransposons have been discovered in plant species. They are generally dispersed over chromosomes, but may also associate with a particular genomic region.

3. 1. rDNA

There are two types of rDNA, 45S rDNA, which contains 18S, 5.8S, 25S rRNA genes together with the transcribed and nontranscribed spacers of preribosomal RNA and 5S rDNA. They are organised in tandem arrays at one or several loci. These two classes of rRNA genes are usually physically separate along the chromosomes of angiosperms but in some species the 5S rDNA is localised on (nucleolar organizing region) NOR-chromosomes, e.g. *Arabidopsis thaliana* (Murata *et al.*, 1997), *Phaseolus vulgaris* (Moscone *et al.*, 1999), *Oryza officinalis* (Shishido *et al.*, 2000), *Brassica campestris* (Hasterok *et al.*, 2001), *Crepis capillaris* (Fig.2a). There is no correlation between the number of 45S and 5S rDNA loci and the number of gene copies in each locus. The 45S and 5S rDNA sites are useful chromosomal landmarks for karyotyping and can provide information about genome evolution and chromosome rearrangements (Fig 2b-d; Hasterok *et al.*, 2001).

The rRNA genes have been intensively investigated in the past decade. Since the adoption of FISH, the knowledge on localization of rDNA on chromosomes and interphase nuclei has significantly increased. The development and improvement of hybridization and detection methods has provided new possibilities to detect rDNA sequences in loci other than the NOR region, offering in this way new chromosome markers (Mukai *et al.*, 1991; Leitch and Heslop-Harrison, 1992; Xu and Earle, 1994). For example, the rDNA sequences can mark B chromosomes (Maluszynska and Schweizer, 1989; Jones 1995) or be present in pericentromeric and terminal regions of *Brassica* chromosomes (Hasterok and Maluszynska, 2000b).

Simultaneous *in situ* hybridization of two rDNA probes allowed for the identification of each chromosome in several species e.g. *Arabidopsis thaliana* (Murata *et al.*, 1997; Fransz *et al.*, 1998) and *Hordeum vulgare* (Leitch and Heslop-Harrison, 1993). FISH with heterologous 18S and 5S

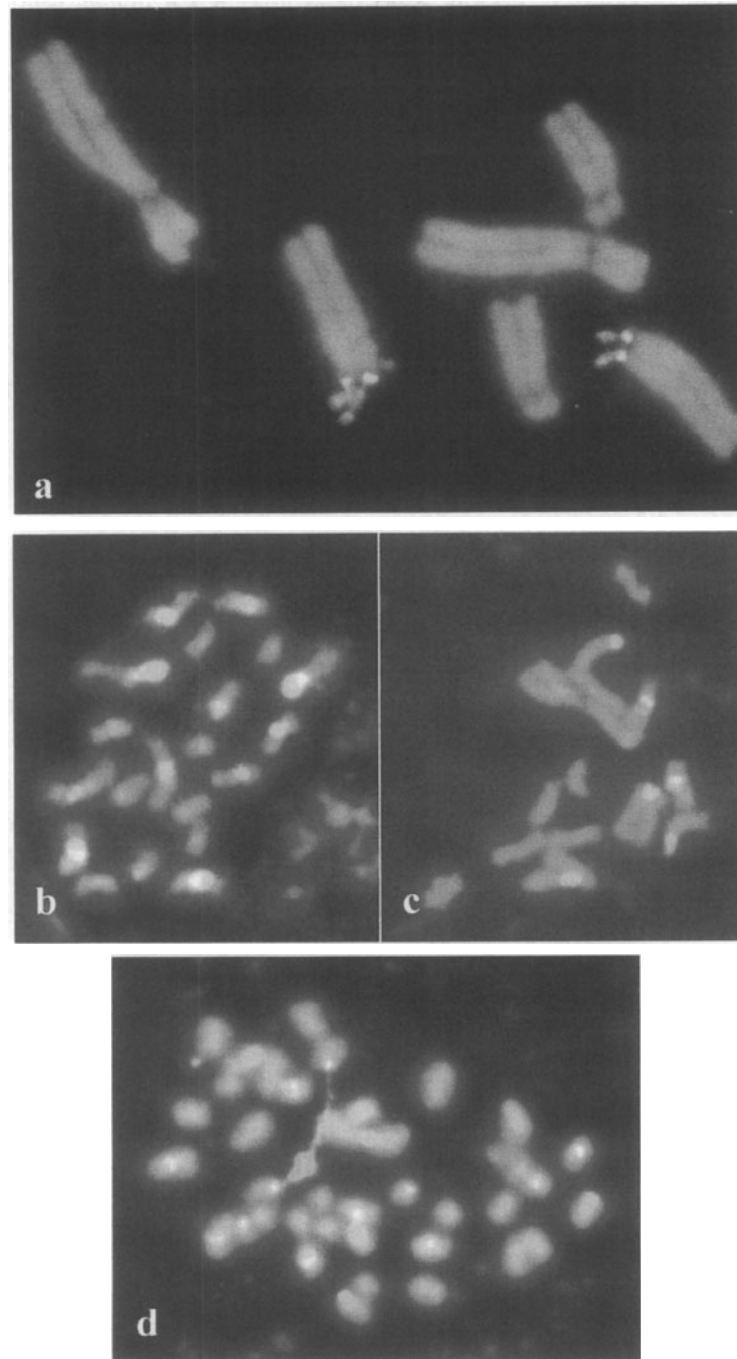


Figure 2. Localization of rRNA genes by double fluorescence *in situ* hybridization of 5S rDNA and 25S rDNA to somatic metaphase chromosomes.

(a) *Crepis capillaris* ($2n = 6$), (b) *Brassica campestris* ($2n=2x=20$; genome AA), (c) *Brassica oleracea* ($2n = 2x = 18$, genome CC), d. *Brassica napus* ($2n = 4x = 38$, genome, AACC) (see Hasterok *et al.*, 2001).

Red (Rodamine) - 5S rDNA; green (FITC) - 25S rDNA; blue - counter-stained with DAPI .

rDNA to barley chromosomes allowed for the identification and orientation of six out of seven chromosome pairs. FISH proved the chromosomal localization of the rDNA signals to trisomic lines of barley. The utility of these FISH landmarks for physical mapping was demonstrated by simultaneous ISH of a BAC clone or cDNA clone (α -amylase-2) with rDNA probes which can mark a particular chromosome or chromosome arm (Brown *et al.*, 1999).

Double-target FISH with 5S and 25S rDNA enables the discrimination of a substantial number of chromosomes from the complement of *Brassica* species. It is possible to distinguish eight types of rDNA distribution in chromosomes of six species of “U-triangle” (Hasterok *et al.*, 2001). For example, *B. campestris* (2n=20) has 12 chromosomes bearing rDNA that represent four types of rRNA gene distribution. The ability to distinguish particular Brassica chromosomes has utility in research with allotetraploids, where the majority of marked chromosomes can be assigned to one or the other of their ancestral genomes, e.g. which chromosomes belong to the AA genome (*B. campestris*) and which ones to the CC genome (*B. oleracea*) in *B. napus* karyotype. (Fig. 2b-d).

In species with several 18S-5.8S-26S rDNA loci only some of them are active. Wheat (*Triticum aestivum*) variety ‘Chinese Spring’ has ten rDNA loci (Mukai *et al.*, 1991) and only those on chromosome 6B and 1B are active in nucleolus formation (Martini and Flavell, 1985), with others being methylated or silent (Flavell *et al.*, 1988). *Brassica campestris* has ten 45S rDNA sites but only two active, and *B. napus* contains four active sites out of 14 45S rDNA loci (Hasterok and Maluszynska, 2000b). Additional chromosome markers can be obtained by silver staining, a cytogenetic method for detecting the position of NORs which were functionally active during the preceding interphase on metaphase chromosomes (Schwarzacher *et al.*, 1978). Sequential application of silver staining and FISH make it possible to distinguish transcriptionally active and inactive rDNA loci in the same metaphase plate (Hasterok and Maluszynska, 2000a; Morais-Cecilio *et al.*, 2000).

Considering rDNA loci as very convenient chromosome markers, it should be noted that the chromosomal pattern of 45S and 5S rDNA could be variable not only between but also within species or even accessions. The variability involves variation in size, number and chromosomal distribution. Moreover, the variation manner of 45S rDNA and 5S rDNA sequences is not parallel. In *A. thaliana* two loci of 45S rDNA have always been detected in all analysed accessions and assigned to short arms of chromosome No. 2 and 4 (Maluszynska and Heslop-Harrison, 1991; Murata *et al.* 1997; Fransz *et al.* 1998), while the localization of 5S rDNA revealed the ecotype-specific polymorphism, concerning both the loci number and their location. In some ecotypes (e.g. Columbia), three 5S rDNA loci were observed on chromosomes No.3, 4 and 5, while in others ecotypes only two

loci were located on chromosomes 4 and 5 (e.g. Wilna). Those two loci are assigned as major ones (Fransz *et al.*, 1998; Weiss and Maluszynska 2000).

Comparison of the localization of rDNA sites on chromosomes of *Oryza* species has shown the lack of variability in the number of 5S rDNA loci per genome, but the sites were located on different chromosomes. Species with genome AA (*O. sativa*), BB (*O. punctata*), and CC (*O. officinalis*) have a 5S rDNA locus on chromosome 11 and species *O. australiensis* (genome EE) and *O. brachyantha* (FF) have such a locus on chromosome 7. In contrast to 5S rDNA, the 45S rDNA loci, occurring on chromosomes 4, 10 and 11 in the genus *Oryza*, exhibit variation in the loci number, as well as position and copy number per locus (Fukui *et al.*, 1994; Shishido *et al.*, 2000). Polymorphism in loci number and their location between varieties of other species has also been reported for genus of *Brassica* (Hasterok *et al.*, 2001), *Phaselous* (Moscone *et al.*, 1999), *Trifolium* (Ansari *et al.*, 1999) and others. Different mechanisms have been postulated as explanations of this phenomenon, i.e. chromosome rearrangement, unequal crossing-over, gene conversion and transpositional events (Schubert and Wobus, 1985; Leitch and Heslop-Harrison, 1993; Hall and Parker, 1995). Comparative analysis of rDNA loci in rice, maize and wheat showed that variation in the location of rDNA sites was in contrast to the high degree of conservation of the order of molecular markers among species. Shishido *et al.* (2000) concluded that the conservation of gene order throughout Gramineae does not apply to the rDNA loci. The position of the rDNA locus could be changed in the genome without chromosome rearrangement during evolution. The mobile nature of rDNA loci has already been proposed for genus *Allium* (Schubert and Wobus, 1985).

3.2. DNA repeats families

Repetitive DNA sequences organised in tandem arrays, known as satellite DNA, can account for a significant percentage of nuclear genomes. They are very useful for studying the taxonomic problems and phylogenetic relationships of plant species, and are helpful for chromosome identification or detection chromosomal rearrangements.

Tandem repetitive sequences have been isolated and characterised for many plant species. The length of the repeat is approximately 160-180 or 320-360 bp, similar to the length of mono- and di-nucleosomes (Martinez-Zapater *et al.*, 1986) repeat. *In situ* hybridization has shown predominant localization of these sequences in subtelomeric or centromeric heterochromatin blocks. However, a few intercalary located satellite DNA families have also been reported. Families of tandem repeats can be specific for single chromosome, genome, species, genus or family.

One of the first isolated satellite DNA was the family of repetitive sequences pSc119, derived from cultivated rye (*Secale cereale*) (Bedbrook *et al.*, 1980). These sequences are located at the telomeric region on all seven chromosomes of rye corresponding to location of the major C-bands. The pSc119 sequence family was detected in all species of the tribe Triticeae and related tribes e.g. Avenae (Katsiotis *et al.*, 1997), but was absent in barley chromosomes. The sequence has been widely used for the examination of the genome evolution and differentiation many species (Vershinin *et al.*, 1996; Taketa *et al.* 2000). Analysis of twelve *Aegilops* species showed that pSc119 hybridized with each of their genomes, but the level and pattern of hybridization was different indicating multiple chromosomal rearrangements had occurred during the genome differentiation of these species (Badaeva *et al.*, 1996). Another repeat DNA family isolated from rye (pSc200) hybridizes at subtelomeric regions of all *S. cereale* standard (A) chromosomes (Vershinin *et al.*, 1995), but is absent from B chromosomes. B chromosomes of rye contain a specific sequence (D1100) located at a subtelomeric heterochromatin region (Langdon *et al.* 2000) (Fig 3 a).

Several different satellite DNA families have been isolated from species of *Beta* genus. They also show variation in distribution, abundance, and chromosomal localization between species. Some satellite DNA repeats can be used as species-specific probes e.g. the satellite family located at centromeres of all chromosomes of sugar beet (*B. vulgaris*). Other satellite DNA repeats are located intercalary and indicate a chromosome-specific variation in repeated array size (Kubis *et al.*, 1998; Schmidt *et al.*, 1998). A repeated sequence specific for A genome of rice has been cloned and localised at the end of the long arm of six pairs of *Oryza sativa* chromosomes (Ohmido and Fukui, 1997). Another repetitive DNA sequence, isolated from rice, was chromosome-specific and has been mapped by FISH to a centromeric heterochromatin region of the long arm of chromosome 5 (Wang *et al.*, 1995). Tandemly repeated DNA sequences isolated from *Zea mays* were located by FISH to a subtelomeric region of chromosomes 3 and 6, and on satellite of the chromosome 6 giving markers for two pairs of maize chromosomes (Chen *et al.*, 2000). Two new families of repetitive sequences, located in different regions of subtelomeric heterochromatin, used for FISH together with sequences (rDNA and Alfa –family) with known distribution patterns have helped to identify ten homologous pairs of chromosomes of *Leymus racemosus* (Kishii *et al.*, 1999).

Martinez-Zapater *et al.* (1986) found that the major satellite DNA sequence (pAL1), isolated from *A. thaliana* is species-specific and located in centromeres of all chromosomes. A hybridization site related to centromeres of all chromosomes is a good indicator of chromosome number even in the interphase nuclei and can be useful for estimation of polyploidization level (Maluszynska and Heslop-Harrison, 1991). More detailed analysis of the *Arabidopsis* genome shows that a tandemly arranged 180bp repeat family is

not the only component of paracentromeric heterochromatin, but there are also some other repetitive sequences present such as telomere-like repeated sequences, *Athila* retroelements, 5S rDNA, mitochondrial DNA, and others (Brandes *et al.*, 1997; Lin *et al.*, 1999).

Tandemly repeated DNA consisting of very short (2-5) nucleotide motifs, known as microsatellites, represents a substantial fraction of the plant genome. Microsatellites provide highly informative and polymorphic markers for gene mapping and genome analysis. Chromosomal distribution of several types of repeats such as (GATA)₄, (GGAT)₄, (CAC)₅, (GA)₈, (CA)₈ and (TA)₁₀ was analysed in sugar beet genome (Schmidt and Heslop-Harrison, 1996). It was indicated that microsatellite sequences showing chromosome-specific amplification have a characteristic ISH pattern. The results have established not only new chromosome markers, but have also shown that amplification and distribution of each microsatellite motif was independent.

The best markers for chromosome identification are chromosome- or species-specific sequences, which have diverged rapidly and extensively during evolution of the species. Nevertheless, even conserved sequences, like telomeric DNA, present in most plant species, can be useful in chromosome identification. Telomeric repeats of sequence motif (TTTAGGG) form an array at the physical end of the chromosomes (Richards and Ausubel, 1988). The array length can vary between individual chromosome arms (Schwarzacher and Heslop-Harrison, 1991; Fuchs and Schubert, 1998). These differences in length can also be useful as chromosome markers and are especially good when they have an intercalary position in a chromosome arm eg. *A. thaliana* (Richards *et al.*, 1991) and *Pinus elliotii* (Doudrick *et al.*, 1995).

3.3. Retrotransposons

Retrotransposons are mobile genetic elements that transpose through reverse transcription of an intermediate RNA. They are present in high-copy numbers in most plants, in some cases comprising over 50% of the genome. Two classes of retrotransposons have been distinguished in plants, LTR retrotransposon (flanked by long terminal repeats) and non-LTR. LTR retrotransposons were subclassified into the TY1-*copia* and the Ty3-*gypsy* groups. Non-LTR retrotransposons have been distinguished into LINEs (long interspersed nuclear elements) and SINEs (short interspersed nuclear elements). Retrotransposons play a major role in plant gene and genome evolution (for review see Kumar and Bennetzen, 1999). Many investigations were focused, over recent years, on the chromosomal organisation of retrotransposons in plant genomes. The distribution pattern of retrotransposons has been analysed, using FISH, in a number of plants,

including Triticeae (Moore *et al.*, 1991; Suoniemi *et al.*, 1996; Pearce *et al.*, 1997; Miller *et al.*, 1998; Langdon *et al.*, 2000), *Beta* (Schmidt *et al.*, 1995), *Gossypium* (Hanson *et al.*, 2000); *Vicia*, *Alstroemeria* (Pearce *et al.*, 1996a; Kuipers *et al.*, 1998) and other species (Brandes *et al.*, 1997). They are generally dispersed over plant chromosomes, but may also be associated with particular chromosome regions. Many retrotransposons can be present in low number in centromeres, interstitial and terminal heterochromatin regions, and rDNA sites (Brandes *et al.*, 1997; Heslop-Harrison *et al.*, 1997; Pearce *et al.*, 1997). However, there are exceptions from this general observation. They are more abundant around centromeres of *Arabidopsis* chromosomes (Brandes *et al.*, 1997; Lin *et al.*, 1999;) or at terminal heterochromatic regions of *Allium cepa* chromosomes (Pearce *et al.*, 1996b). There are several reports on the presence of retrotransposon sequences in centromeric regions of cereal species including barley, wheat and rye (Aragon-Alcaide *et al.*, 1996; Presting *et al.*, 1998; Langdon *et al.*, 2000). These sequences can also be useful for labelling by FISH the centromere regions of chromosomes.

3.4. Gene mapping

Gene mapping by ISH and cytogenetic map construction of chromosomes enables the integration of genetic and physical maps. Cytogenetic maps display the approximate positions of genes in relation to structural chromosome markers such as centromere, telomeres and heterochromatin blocks. Distance on a genetic map is based on the frequency of meiotic recombinations between genes or markers. The order of genes along the chromosome on genetic, physical and cytogenetic maps is of course the same, but distances can differ significantly. Two genes that are distantly located on the genetic map may be physically quite close, whereas closely linked genes may be physically separated by a large distance (Gustafson *et al.*, 1990; Heslop-Harrison, 1991; Leitch *et al.*, 1994; Jiang *et al.*, 1996). The ratio of the genetic distance (cM) to physical distance (kb) can be 1cM per 10 – 100 000 kb depending on the species and the particular chromosome region. In tomato (*Lycopersicum esculentum* Mill.) chromosomes the ratio is 1cM/750kb on average, but the centromeric region of chromosome No. 9 it is 1 cM/ 4Mb (Frary *et al.*, 1996). This is the result of different crossing-over frequencies along chromosome. The recombination level is particularly low at the centromere and heterochromatin regions. The comparison of the genetic and physical maps of wheat and barley shows that recombinations mainly occur in the distal part of chromosomes (Pedersen *et al.*, 1995; Gill *et al.*, 1993). Cytogenetic maps of plant chromosomes are lagging behind genetic and physical maps. Only single genes were mapped on chromosomes by ISH, e.g. *Sec-1* in rye (Gustafson *et al.*, 1990; Busch *et al.*, 1994; Clarke *et al.*, 1996), B-hordein (Lehler *et al.*,

1993), α -amylase-2 in barley (Leitch and Heslop-Harrison, 1993), *sh2* in sorghum (Gomez *et al.*, 1997), and chalcone synthase (*chsA*) in petunia (Fransz, *et al.*, 1996b).

Very important for gene mapping is the ability to locate two or more DNA probes simultaneously. In plant genome investigations FISH with two DNA probes, double FISH, is routinely used for physical chromosome mapping (Leitch *et al.*, 1991a; Maluszynska and Heslop-Harrison, 1991; Lim *et al.*, 2000). Several DNA sequences can be ordered in relation to each other by applying multicolour FISH (M-FISH). Many different fluorescence colours of hybridization sites can be achieved by combinatorial labelling of probes with defined combination of fluorophores (Mukui, *et al.* 1993b). It is possible to have seven combinations of labelling with the use of 3 fluorophores. Changing the ratio of fluorophores can increase the number of possible combinations. In human cytogenetics, M-FISH resulting in 24 colours is applied in basic studies and diagnostic (Lichter, 1997).

FISH localization of single-copy plant genes has some limitations mainly dealing with the resolution and sensitivity of the detection. The mapping resolution is defined as the smallest physical distance between target sequences that can be resolved with a fluorescence microscope. The distance between sequences depends on the chromosome stage and their condensation. The resolution on metaphase chromosomes is about 2-5 Mb while interphase nuclei provide better resolution, between 50-100 kb. Sensitivity, which means the smallest length of the detecting sequence, is on the order of 1 to 2 kb for plant metaphase chromosomes (de Jong *et al.*, 1999).

Recently, new high-resolution FISH techniques have been developed with application of pachytene chromosomes or extended DNA fibres (EDF). Another strategy to increase the detection sensitivity is to use, as a probe for FISH, a large insert of DNA clones in BACs or YACs (de Jong *et al.*, 1999). These techniques are especially suitable for plants with small genomes. Genes and YACs were localized successfully on pachytene chromosomes of such plants as *Lycopersicum*, *Petunia*, and *Arabidopsis* (Zhong, *et al.*, 1996; 1998; de Jong *et al.*, 1999). Better resolution can be obtained by applying FISH to extended DNA fibres from lysed nuclei. Hybridization signals on the linearized chromosomal DNA reflect the physical position of the probes along the DNA. A set of BAC clones, specific for each of the 12 potato (*Solanum tuberosum* L.) chromosomes was used to mark small chromosomes of this species (Dong *et al.*, 2000).

The order of clones, their distance and the size of DNA sequences was determined using multicolour FISH to *Arabidopsis* and tomato extended DNA fibres (Fransz *et al.*, 1996a; Zhong *et al.*, 1996; Mesbah *et al.*, 2000). The FISH to extend DNA showed a difference in copy number of several repetitive sequences between two cultivars of *indica* and *japonica* rice (Ohmido *et al.*, 2000). This technique allowed localization of T-DNA constructs in transgenic potato plants. Differently labelled probes of T-DNA and vector sequences

made it possible to determine the copy number and arrangement of T-DNA in potato chromosomes (Wolters *et al.*, 1998).

Transgenic plants are widely used in fundamental studies of gene expression and agricultural biotechnology. The number of transgene copies and the position of the integration into plant chromosomes are crucial for their expression (Iglesias *et al.*, 1997). Transgene localization by FISH has been reported in transgenic plants obtained through T-DNA mediated transfer (Ambros *et al.*, 1986; Moscone *et al.*, 1996; ten Hoopen *et al.* 1996; 1999) and via particle bombardment (Pedersen *et al.*, 1997; Svitashv *et al.*, 2000). Cytogenetic analysis revealed uneven distribution of transgenes along chromosomes and the tendency to integrate toward distal chromosome regions (ten Hoopen *et al.* 1999; Pedersen *et al.*, 1997).

4. IDENTIFICATION AND DISCRIMINATION OF GENOMES BY GENOMIC *IN SITU* HYBRIDIZATION

Genomic *in situ* hybridization (GISH) involves the extraction of total genomic DNA from one species, labelling and using as probe for hybridization to target chromosomes or interphase nuclei of related species. GISH has been successfully applied to identify parental genomes in hybrids (Ananthawat-Jonsson *et al.*, 1990; Schwarzacher *et al.*, 1992), allopolyploids (Bennett *et al.*, 1992), and for detection of alien chromosomes or chromosome fragments (Heslop-Harrison *et al.*, 1990). GISH was also used for homologous or homeologous chromosome pairing analysis (Le and Armstrong 1991; King *et al.*, 1993; Martin *et al.*, 1998). *In situ* hybridization of genomic DNA of one species to genomic DNA of related species can provide information about phylogenetic relationships and chromosomal rearrangements during the evolution of these species (Orgaard and Heslop-Harrison, 1994).

Genomic nuclear DNA extracted from the species of interest is cut into small fragments (500 –1000 bp) and labelled using standard protocols. Generally, chromosome preparation and conditions for denaturation, hybridization and detection are similar to that described for FISH. The addition of an excess of unlabeled, sheared DNA from the parent species not used as a probe – "blocker DNA" - is recommended for increasing the specificity of hybridization (Heslop-Harrison and Schwarzacher, 1996). The numerous protocols for GISH are described in detail in recent publications and manuals.

4.1. Genome discrimination

It has been estimated that the majority of angiosperms are polyploids, mostly allopolyploids, which includes most crop plants (Leitch and Bennett, 1997;

Soltis and Soltis, 2000). Many plants previously considered to be a diploid are actually stabilized or chromosomally diploidized polyploids as demonstrated by either comparative chromosome mapping or meiotic chromosome behaviour e.g. maize (Gaut *et al.*, 2000). Taxonomic relationships between maize and several of its wild relatives, investigated by GISH, provided new interesting results, but did not solve the problem of maize ancestral diploid species (Takahashi *et al.*, 1999).

Evidence for polyploid origin of many plants has been obtained through comparative genomics based on restriction fragment length polymorphism (RFLP) and chromosome mapping. Recently developed GISH provides new possibilities for ancestral species determination. GISH has been used to test a phylogeny of different allopolyploid plant species such as *Milium montianum* (Bennett, *et al.*, 1992), *Allium cepa* var. *viviparum* (Puzina *et al.*, 1999), garden cultivars of *Crocus* (Orgaard *et al.*, 1995), *Dalia* species (Gatt, *et al.*, 1999), and cultivated *Musa* spp. (Osuji *et al.*, 1997; D'Hont *et al.*, 2000). GISH was successfully used for discrimination of genomes in a wide range of Triticeae species (Ananthawat-Jonsson, *et al.*, 1990), and in interspecific hybrids of even close related species as *H. vulgare* x *H. bulbosum* (Ananthawat-Jonsson, *et al.*, 1993). Many parental genomes have been discriminated in hybrids including species involving the genera *Triticum*, *Hordeum*, *Secale*, *Aegilops*, *Elymus*, *Leymus* and others (Kosina and Heslop-Harrison, 1993; Lima-Brito *et al.*, 1997; Chen *et al.*, 1998). The total genomic DNA of rye allowed the labelling of rye chromosomes in triticale (X *Triticosecale* Wittmuck) genome (Fig. 3c, d). The rice chromosomes of tetraploid species *O. minuta* (BBCC) and *O. latifolia* (CCDD) have been identified by GISH with total genomic DNA of *O. officinalis* (CC) (Fukui *et al.*, 1997). It is also possible to use differentially labelled total genomic DNA of both parents and perform simultaneous GISH to chromosomes of a hybrid, as demonstrated with Triticeae hybrids (Ananthawat-Jonsson and Reader, 1995; Kosina and Heslop-Harrison, 1996). Mukai *et al.* (1993b) used mFISH to distinguish all three genomes of hexaploid wheat.

GISH allows not only identification of parental or ancestral genomes, but also the detection of chromosomal rearrangements. The analysis of tobacco (*Nicotiana tabacum*) genome structure by GISH was among the earliest reported studies (Kenton *et al.*, 1993). It was confirmed that tobacco is an allotetraploid whose parents are *N. sylvestris* (S genome) and a species with T-genome from the section *Tomentosae*. The analysis revealed that numerous chromosomal rearrangements took place between chromosomes originating from these two ancestral species. Intergenomic translocations were also detected by GISH in other species such as allotetraploid *Avena maroccana*, allohexaploid *A. sativa* (Jellen *et al.*, 1994; Chen and Armstrong 1994), and *A. fatua* (Yang *et al.*, 1998). GISH was also used to detect chromosomes of raspberry (*Rubus idaeus*) in the genome of allopolyploid plants of blackberry (*R. aggregate*) and revealed translocations between them

(Lim *et al.*, 1998). The application of GISH to the parental species, F_1 and backcross hybrid plants enabled the discrimination of both genomes. Moreover, it provided, in the backcross, visual evidence of the cross-over site position that occurred in the F_1 meiosis. For example, the recombination between parental genomes was found in backcross of bigeneric hybrids *Gasteria lutzii* x *Aloe arista* (Takahashi *et al.*, 1997), and *Lolium multiflorum* x *Festuca pratensis* (Thomas *et al.*, 1994).

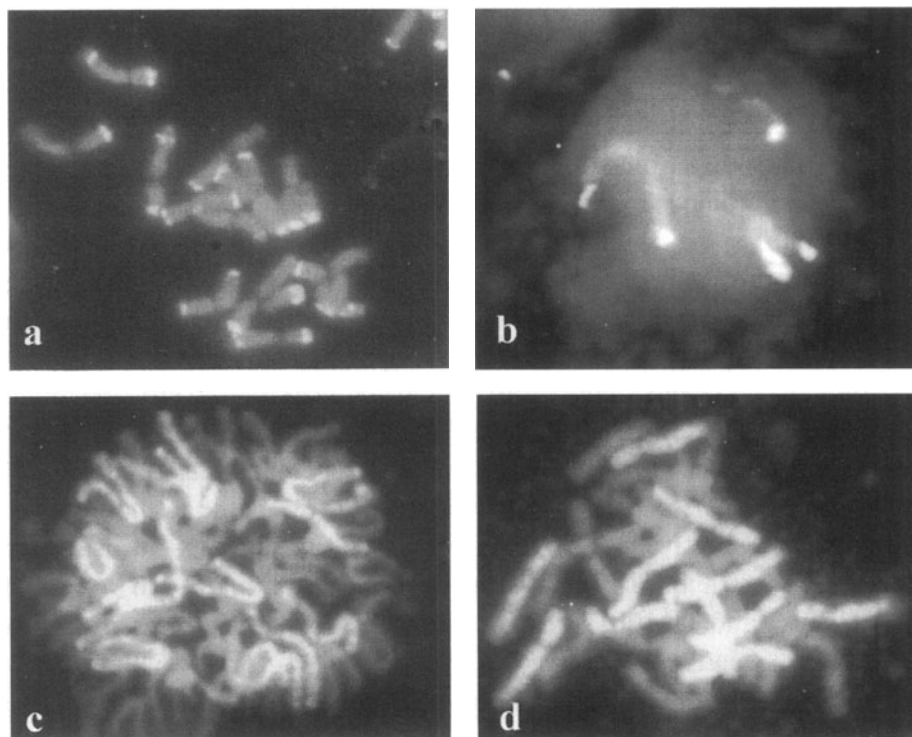


Figure 3. (a) *In situ* hybridization of repetitive DNA sequences to metaphase chromosome of *Secale cereale* Exp. B ($2n=14+2B$).

Red (Rodamine) – D1100 B rye specific DNA sequence; green (FITC) – pSc200 sequence hybridize to all A chromosomes; blue - counter-stained with DAPI

(b) GISH and FISH to interphase nucleus of *Triticum aestivum* cv. Lindstrom with alien B chromosomes of rye ($2n = 6x = 42 = 4B$).

Red (Rodamine) – D1100 B rye specific DNA sequence; green/white (FITC) – genomic DNA of rye without B, blue - counter-stained with DAPI

(c, d) Genome discrimination by *in situ* hybridization. in prometaphase and metaphase cells of triticale cv. Lasko ($2n = 6x = 42$, genome AABBRR).

Rodamine labelled genomic DNA of rye hybridized with chromosomes belonging to rye genome (red), wheat chromosomes are blue after DAPI staining.

(Kindly provided by Dr. R. Hasterok)

4.2. Identification of alien chromosomes and chromosome segments

In plant breeding programmes translocation, addition, recombinant and substitution lines are a way to enhance genetic variability for crop improvement. Translocations are the most extensively studied chromosome rearrangements in the tribe of Triticeae. They are involved in the transfer of desirable genes from alien species to crops such as wheat. Using this approach, several genes of resistance to fungal and viral diseases, to pests, and to abiotic stresses have been transferred from various *Aegilops*, *Agropyron*, *Secale*, and other species (Friebe *et al.*, 1991; 1993; Mukai *et al.*, 1993a) into wheat. Plants carrying translocated chromosomes can be identified at meiosis, but this requires plant cultivation and time consuming meiotic analyses. *In situ* hybridization provides the possibility for early detection of translocations at the seedling stage. An intergenomic transfer can be visualised by GISH. One of the first uses of fluorescent GISH for detection of translocations has been achieved through application of genomic rye DNA involving 1B/1R translocation chromosomes of hexaploid wheat (Heslop-Harrison *et al.*, 1990). An advantage is the ability to identify alien chromosomes or chromosome segments not only in metaphase, but also in nuclei, at all the cell cycle stages (Leitch *et al.*, 1991). In various wheat addition, substitution and translocation lines, the chromosomes from *H. vulgar*, *H. chilense*, *Thinopyrum bessarabicum*, *Leymus multicaulis*, and *S. cereale* were identified in metaphase (Ribeiro-Carvalho *et al.*, 1997; Chen *et al.*, 1999; Hohmann, *et al.*, 1999; Tang *et al.*, 2000; Friebe *et al.*, 2000; Molnar-Lang *et al.*, 2000). In interphase nuclei one or two labelled domains were visible in monosomic and disomic lines respectively (Schwarzacher *et al.*, 1992). Fig. 3a demonstrates double *in situ* hybridization with genomic DNA of rye (without B chromosomes) and D1100 - rye B-specific DNA sequences as probes to *T. aestivum* cv. Lindstrom (Langdon *et al.*, 1999). The “painting” of B chromosomes with two fluorescence colours allowed detection of rye B chromosomes in wheat interphase nucleus.

More precise identification and localization of chromosome rearrangement can be obtained by the use of combined FISH/GISH, which means GISH with genomic DNA as a probe to distinguish genomes, and FISH with chromosome-specific DNA probes to identify different chromosomes or their segments in hybrid genome (Moscone *et al.*, 1996; Lim *et al.*, 1998; Poggio *et al.*, 1999; D’Hont *et al.*, 2000; Pickering *et al.*, 2000). This approach has been illustrated by the work of Castilho *et al.* (1996) on wheat-*Ae. umbellulata* lines. They applied genomic DNA from *Ae. umbellulata* together with 5S rDNA, 18-15S rDNA, and repeat sequences from rye for characterization of 11 recombinant lines. The result indicated physical location of the breakpoints and the number and size of alien chromosome segments.

4. 3. Chromosome pairing

GISH has also been a very valuable method for analysis of *Solanum tuberosum* and *L. esculentum* hybrids developed by somatic fusion. The use of the total genomic DNA of tomato reveals not only the number of tomato chromosomes in the hybrid, but also visualises the homoeologous pairing and crossing-over between chromosomes of these two species (Escalante *et al.*, 1998; Garriga-Caldere, *et al.*, 1997; 1999). The study of homoeologous chromosome pairing and chiasma formation can provide information on the relationship between genomes of hybrid or amphiploid lines and are important for speeding up the introduction of useful genes from wild species to crops (King *et al.*, 1993). The application of GISH to meiotic chromosomes and the “painting” of the chromosomes of one genome offered greater precision in identifying constituent genomes in a hybrid. GISH made possible the observation of alien chromosome behaviour in wheat meiosis (Fedak *et al.*, 2000). The effect of the *Ph1* locus on the behaviour of rye telosomes in a wheat background was analysed by GISH, using genomic rye DNA as a probe (Mikhailova *et al.*, 1998). Application of GISH to the meiotic chromosomes of the *Lolium-Festuca* complex revealed a close relationship between these species (Cao *et al.*, 2000). Introgression of genes from *Allium fistulosum* into *A. cepa* using *A. roylei* as a bridging species was studied by the use of genomic mFISH to both mitotic and meiotic chromosomes (Khrustaleva and Kik, 2000). This technique was also effective in the characterization of meiotic pairing involving the relatively small chromosomes in *O. sativa* x *O. australiensis* hybrids (Abbasi *et al.*, 1999).

5. CONCLUSIONS

In situ hybridization, discussed here, is the main technique of molecular cytogenetics, which includes flow cytometry, micromanipulation, and recently microarray hybridization. An advance in cytogenetic investigations has followed the intensive development of molecular biology. There are many fundamental, unsolved problems related to genome structure and function, gene mapping, role of repetitive DNA sequences, chromosome behaviour in mitosis and especially meiosis, which can be explained with the help of molecular cytogenetics. The further development of cytogenetics is also necessary for rapid progress in applied research including genetic manipulation for crop improvement.

6. REFERENCES

- Abbasi, F. M., Brar, D. S., Carpena, A. L., Fukui, K., and Khush, G. S. (1999) Detection of autosyndetic and allosyndetic pairing among A and E genomes of *Oryza* through genomic *in situ* hybridization. *Rice Genetic Newsletter* **16**, 24-25.
- Ambros, P. F., Matzke, M. A., and Matzke, J. M. (1986) Detection of 17 kb unique sequence (T-DNA) in plant chromosomes by *in situ* hybridization. *Chromosoma* **94**, 11-18.
- Anamthawat-Jonsson, K. and Reader, S. M. (1995) Pre-annealing of total genomic DNA probes for simultaneous genomic *in situ* hybridization. *Genome* **38**, 814-816.
- Anamthawat-Jonsson, K., Schwarzacher, T., and Heslop-Harrison, J. S. (1993) Behavior of parental genomes in the hybrid *Hordeum vulgare* x *H. bulbosum*. *J. Hered.* **84**, 78-82.
- Anamthawat-Jonsson, K., Schwarzacher, T., Leitch, A. R., Bennett, M. D., and Heslop-Harrison, J. S. (1990) Discrimination between closely related *Triticeae* species using genomic DNA as a probe. *Theor. Appl. Genet.* **79**, 721-728.
- Ansari, H. A., Ellison, N. W., Reader, S. M., Badaeva, E. D., Friebe, B., Miller, T. E., and Williams, W. M. (1999) Molecular cytogenetic organization of 5S and 18S-26S r DNA loci in white clover (*Trifolium repens* L) and related species. *Ann. Bot.* **83**, 199-206.
- Aragon-Alcaide, L., Miller, T., Schwarzacher, T., Reader, S., and Moore, G. (1996) A cereal centromeric sequence. *Chromosoma* **105**, 261-268.
- Badaeva, E. D., Friebe, B., and Gill, B. S. (1996) Genome differentiation in *Aegilops*. 2. Physical mapping of 5S and 18S-26S ribosomal RNA gene families in diploid species. *Genome* **39**, 1150-1158.
- Bedbrook, J. R., Jones, J., O'Dell, M., Thompson, R. D., and Flavell, R. B. (1980) Molecular characterisation of telomeric heterochromatin in *Secale* species. *Cell* **19**, 545-560.
- Bennett, M. D. (1998) Plant genome values: how much do we know? *Proc. Natl. Acad. Sci. USA* **95**, 2011-2016.
- Bennett, S. T., Kenton, A. Y., and Bennett, M. D. (1992) Genomic *in situ* hybridization reveals the allopolyploid nature of *Milium montianum* (Gramineae). *Chromosoma* **101**, 420-424.
- Brandes, A., Heslop-Harrison, J. S., Kamm, A., Kubis, S., Doudrick, R. L., and Schmidt, T. (1997) Comparative analysis of the chromosomal and genomic organization of Ty1-copia-like retrotransposons in pteridophytes, gymnosperms and angiosperms. *Plant Mol. Biol.* **33**, 11-21.
- Brown, S. E., Stephens, J. L., Lapitan, N. L. V., and Knudson, D. L. (1999) FISH landmarks for barley chromosomes (*Hordeum vulgare* L.). *Genome* **42**, 274-281.
- Busch, W., Herrmann, R. G., and Martin, R. (1995). Refined physical mapping of the *Sec-1* locus on the satellite of chromosome 1R of rye (*Secale cereale*). *Genome* **38**, 889-893.
- Cao, M., Sleper, D. A., Dong, F., and Jiang, J. (2000) Genomic *in situ* hybridization (GISH) reveals high chromosome pairing affinity between *Lolium perenne* and *Festuca mairei*. *Genome* **43**, 398-403.
- Caspersson, T., Farber, S., Foley, G. E., Kudynowski, E. J., Modest, E. J., Simonsson, E., Wagh, U., and Zech, L. (1968) Chemical differentiation along metaphase chromosomes. *Exp. Cell Res.* **49**, 214-222.
- Castilho, A., Miller, T. E., and Heslop-Harrison, J. S. (1996) Physical mapping of translocation breakpoints in a set of wheat-*Aegilops umbellulata* recombinant lines using *in situ* hybridization. *Theor. Appl. Genet.* **93**, 816-825.
- Chen, C. C., Yan, H., Zhai, W., Zhu, L., and Sun, J. (2000) Identification and chromosomal location of a new tandemly repeated DNA in maize. *Genome* **43**, 181-184.
- Chen, Q. and Armstrong, K. (1994) Genomic *in situ* hybridization in *Avena sativa*. *Genome* **37**, 607-612.

- Chen, Q., Conner, R. L., Laroche, A., Ji, W., Armstrong, K. C., and Fedak, G. (1999) Genomic *in situ* hybridization analysis of *Thinopyrum* chromatin in a wheat *Th. intermedium* partial amphiploid and six derived chromosome addition lines. *Genome* **42**, 1217-1223.
- Chen, Q., Conner, R. L., Laroche, A., and Thomas, J. B. (1998) Genome analysis of *Thinopyrum intermedium* and *Th. ponticum* using genomic *in situ* hybridization. *Genome* **41**, 580-586.
- Clarke, B. C., Mukai, Y., and Appels, R. (1996) The *Sec-1* locus on the short arm of chromosome 1R of rye (*Secale cereale*). *Chromosoma* **105**, 269-275.
- D'Hont, A., Paget-Goy, A., Escoute, J., and Carreel, F. (2000) The interspecific genome structure of cultivated banana, *Musa spp.* revealed by genomic DNA *in situ* hybridization. *Theor. Appl. Genet.* **100**, 183.
- Darlington, C. D. (1937) "Recent Advances in Cytology," Churchill, London.
- deJong, J. H., Fransz, P., and Zabel, P. (1999). High-resolution FISH in plants- techniques and applications. *Trends in Plant Sci.* **4**, 258-263.
- Dong, F., Song, J., Naess, S. K., Helgeson, J. P., Gebhardt, C., and Jiang, J. (2000) Development and application of a set of chromosome-specific cytogenetic DNA markers in potato. *Theor. Appl. Genet.* **101**, 1001-1007.
- Doudrick, R. L., Heslop-Harrison, J. S., Nelson, C. D., Schmidt, T., Nance, W. L., and Schwarzach, T. (1995) Karyotype of slash pine (*Pinus elliottii* var. *elliottii*) using patterns of fluorescence *in situ* hybridization and fluorochrome banding. *J Hered* **86**, 286-296.
- Escalante, A., Imanishi, S., Hossain, M., Ohmido, N., and Fukui, K. (1998) RFLP analysis and genomic *in situ* hybridization (GISH) in somatic hybrids and their progeny between *Lycopersicon esculentum* and *Solanum lycopersicoides*. *Theor. Appl. Genet.* **96**, 719-726.
- Fedak, G., Chen, Q., Conner, R. L., Laroche, A., Petroski, R., and Armstrong, K. W. (2000) Characterization of wheat-*Thinopyrum* partial amphidiploids by meiotic analysis and genomic *in situ* hybridization. *Genome* **43**, 712-719.
- Flavell, R. B., O'Dell, M., and Thompson, W. F. (1988) Regulation of cytosine methylation in ribosomal DNA and nucleolus organizer expression in wheat. *J. Mol. Biol.* **204**, 523-534.
- Fransz, P. F., Alfonso-Blanco, C., Liharska, T. B., Peeters, A. J. M., Zabel, P., and de Jong, H. J. (1996a) High-resolution physical mapping in *Arabidopsis thaliana* and tomato by fluorescence *in situ* hybridization to extended DNA fibers. *Plant J.* **9**, 421-430.
- Fransz, P. F., Stam, M., ten Hoopen, R., and Nanninga, N. (1996b) Detection of single-copy genes and chromosome rearrangements in petunia hybrids by fluorescence *in situ* hybridization. *Plant J.* **9**, 767-774.
- Fransz, P. F., Armstrong, K., Alonso-Blanco, C., Fisher, T. C., Trres-Ruiz, R., and Jones, G. (1998) Cytogenetics for model system *Arabidopsis thaliana*. *Plant J.* **13**, 421-430.
- Frary, A., Presting, G. G., and Tanksley, S. D. (1996) Molecular mapping of the centromeres of tomato chromosomes 7 and 9. *Mol. Gen. Genet.* **250**, 295-304.
- Friebe, B., Jiang, J., Gill, B. S., and Dyck, P. L. (1993) Radiation-induced nonhomoeologous wheat-*Agropyron intermedium* chromosomal translocations conferring resistance to leaf rust. *Theor. Appl. Genet.* **86**, 141-149.
- Friebe, B., Kynast, R. G., and Gill, B. S. (2000) Gametocidal factor-induced structural rearrangements in rye chromosomes added to common wheat. *Chromosome Res.* **8**, 501-511.
- Friebe, B., Mukai, Y., Dhaliwal, H. S., Martin, T. J., and Gill, B. S. (1991) Identification of alien chromatin specifying resistance to wheat streak mosaic and greenbug in wheat germplasm by C-banding and *in situ* hybridization. *Theor. Appl. Genet.* **81**, 381-389.

- Fuchs, J. and Schubert, I. (1998) Characterization of plant genomes using fluorescence *in situ* hybridization. In "Plant Cytogenetics" (J. Maluszynska, Ed.), Wydawnictwo Uniwersytetu Slaskiego, Katowice.
- Fukui, K. And Nakayama, S. Eds (1996) Plant Chromosomes: Laboratory Methods. Boca Raton: CRC Press.
- Fukui, K., Ohmido, N., and Khush, G. S. (1994) Variability in rDNA loci in the genus *Oryza* detected through fluorescence *in situ* hybridization. *Theor. Appl. Genet.* **87**, 893-899.
- Fukui, K., Shishido, R., and Kinoshita, T. (1997) Identification of the rice D-genome chromosomes by genomic *in situ* hybridization. *Theor. Appl. Genet.* **95**, 1239-1245.
- Garriga-Caldera, F., Huigen, D. J., Filotico, F., Jacobsen, E., and Ramanna, M. S. (1997) Identification of alien chromosomes through GISH and RFLP analysis and the potential for establishing potato lines with monosomic additions of tomato chromosomes. *Genome* **40**, 666-673.
- Garriga-Caldera, F., Huigen, D. J., Jacobsen, E., and Ramanna, M. S. (1999) Prospects for introgressing tomato chromosomes into the potato genome: An assessment through GISH analysis. *Genome* **42**, 282-288.
- Gatt, M., Hammett, K., Murray B. (1999) Confirmation of ancient polyploidy in *Dahlia* (Asteraceae) species using genomic *in situ* hybridization. *Ann. Bot.* **84**: 39-48.
- Gaut, B. S., Le Thierry d'Ennequin, M., Peek, A. S., and Sawkins, M. C. (2000) Maize as a model for the evolution of plant nuclear genomes. *Proc. Natl. Acad. Sci. USA* **97**, 7008-7015.
- Gill, K. S., Gill, B. S., and Endo, T. R. (1993) A chromosome region-specific mapping strategy reveals gene-rich telomeric ends in wheat. *Chromosoma* **102**, 374-381.
- Gomez, M. I., Islam-Faridi, M. N., Woo, S.-S., Schertz, K. F., Czeschin, D., Zwick, M. S., Wing, R. A., Stelly, D. M., and Price, J. H. (1997) FISH of a maize *sh2*-selected sorghum BAC to chromosomes of *Sorghum bicolor*. *Genome* **40**, 475-478.
- Gustafson, J. P. (1983) Cytogenetics of triticale. In "Cytogenetics of Crop Plant" (M. S. Swaminathan, P. K. Gupta, and U. Sinha, Eds.), MacMillan India, Delhi.
- Gustafson, J. P., Butler, E., and McIntyre, C. L. (1990). Physical mapping of low-copy DNA sequences in rye (*Secale cereale* L.). *Proc. Natl. Acad. Sci. USA* **87**, 1899-1902.
- Hall, K. J. and Parker, J. S. (1995) Stable chromosome fission associated with rDNA mobility. *Chromosome Res.* **3**, 417-422.
- Hanson, R. E., Islam-Faridi, M. N., Crane, C. F., Zwick, M. S., Czeschin, D. G., Wendel, J. F., McKnight, T. D., Price, H. J., and Stelly, D. M. (2000) Ty1-*copia*-retrotransposon behaviour in a polyploid cotton. *Chromosome Res.* **8**, 73-76.
- Hasterok, R., Jenkins, G., Langdon, T., Jones, N., and Maluszynska, J. (2001) Ribosomal DNA is an effective marker of Brassica chromosomes. *Theor. Appl. Genet.* **103**:486-490.
- Hasterok, R. and Maluszynska, J. (2000a) Different rRNA gene expression in primary and adventitious roots of *Allium cepa* L. *Folia Histochemica et Cytobiologica* **38**, 181-184.
- Hasterok, R. and Maluszynska, J. (2000b) Nucleolar dominance does not occur in root tip cells of allotetraploid *Brassica* species. *Genome* **43**, 574-579.
- Heslop-Harrison, J. S. (1991) The molecular cytogenetics of plant. *J. Cell Sci.* **100**, 15-21.
- Heslop-Harrison, J. S., Brandes, A., Taketa, S., Schmidt, T., Vershinin, A. V., Alkhimova, E. G., Kamm, A., Doudrick, R. L., Schwarzacher, T., Katsiotis, A., Kubis, S., Pearce, S. R., Flavell, A. J., and Harrison, G. E. (1997) The chromosomal distribution of Ty1-*copia* group of retrotransposable elements in higher plants and their implication for genome evolution. *Genetica* **100**, 197-204.
- Heslop-Harrison, J. S., Leitch, A. R., Schwarzacher, T., and Ananthawat-Jonsson, K. (1990) Detection and characterization of 1B/1R translocations in hexaploid wheat. *Heredity* **65**, 385-392.

- Heslop-Harrison, J.S. and Schwarzacher, T. (1996) Genomic Southern and *in situ* hybridization for plant genome analysis. In: Jauhar PP, ed. *Methods of Genome Analysis in Plants*. Boca Raton: CRC, 163-179.
- Hohmann, U., Zoller, J., Herrmann, R. G., and Kazman, M. E. (1999) Physical mapping and molecular-cytogenetic analysis of substitutions and translocations involving chromosome 1D in synthetic hexaploid triticale. *Theor. Appl. Genet.* **98**, 647-656.
- Iglesias, V. A., Moscone, E. A., Papp, I., Neuhuber, F., Michalowski, S., Phelan, T., Spiker, S., Matzke, M., and Matzke, A. J. M. (1997) Molecular and cytogenetic analyses of stably and unstably expressed transgene loci in tobacco. *Plant Cell* **9**, 1251-1264.
- Jellen, E. N., Gill, B. S., and Cox, T. S. (1994) Genomic *in situ* hybridization differentiates between A/D and C-genome chromatin and detects intergenomic translocations in poliploid oat species (genus *Avena*). *Genome* **37**, 613-618.
- Jiang, J., Hulbert, S. H., Gill, B. S., and Ward, D. C. (1996) Interphase fluorescence *in situ* hybridization mapping: a physical mapping strategy for plant species with large complex genomes. *Mol. Gen. Genet.* **252**, 497-502.
- Jones, R. N. (1995) B chromosomes in plants. Transley Review No. 85. *New Phytol.* **131**, 411-434.
- Katsiotis A., Hagidimitriou, M., Heslop-Harrison, J.S. (1997) The close relationship between the A and B genomes in *Avena L.* determined by molecular cytogenetic analysis of total genomic, Tandemly and dispersed repetitive DNA sequences *Ann. Bot.* **79**: 103-109
- Kenton, A., Parokonny, A. S., Gleba, Y. Y., and Bennett, M. D. (1993) Characterization of the *Nicotiana tabacum L.* genome by molecular cytogenetics. *Mol. Gen. Genet.* **240**, 159-169.
- Khrustaleva, L.I., Kik, C. (2000) Introgression of *Allium fistulosum* into *A. Cepa* mediated by *A. roylei*. *Theor. Appl. Genet.* **100**: 17-26.
- King, I. P., Purdie, K. A., Orford, S. E., Reader, S. M., and Miller, T. E. (1993) Detection of homoeologous chiasma formation in *Triticum durum* x *Thinopyrum bessarabicum* hybrids using genomic *in situ* hybridization. *Heredity* **71**, 369-372.
- Kishii, M., Nagaki, K., Tsujimoto, H. and Sasakuma, T. (1999) Exclusive localization of tandem repetitive sequences in subtelomeric heterochromatin regions of *Leymus racemosus* (Poaceae, Triticeae). *Chromosome Res.* **7**, 519-529.
- Kosina, R. And Heslop-Harrison J.S. (1996) Molecular cytogenetics of an amphiploid trigeneric hybrid between *Triticum durum*, *Thinopyrum distichum* and *Lophopyrum elongatum*. *Ann. Bot.* **78**: 583-589
- Kubis, S., Schmidt, T., and Heslop-Harrison, J. S. (1998) Repetitive DNA elements as a major component of plant genomes. *Ann. Bot.* **82**, 45-55.
- Kuipers, A. G. J., Heslop-Harrison, J. S. P., and Jacobsen, E. (1998) Characterisation and physical location of Ty1-*copia*-like retrotransposons in four *Alstroemeria* species. *Genome* **41**, 357-367.
- Kumar, A. and Bennetzen, J. L. (1999) Plant retrotransposons. *Annu. Rev. Genet.* **33**, 479-532.
- Langdon, T., Seago, Ch., Jones, R. N., Oughan, H., Thomas, H., Forster, J. W., and Jenkins, G. (1999) *De novo* evolution of satellite DNA on rye B chromosome. *Genetics* **154**, 869-884.
- Langdon, T., Seago, Ch., Mende, M., Leggett, M., Thomas, H., Forster, J. W., Jones, R. N., and Jenkins, G. (2000) Retrotransposon evolution in diverse plant genomes. *Genetics* **156**, 313-325.
- Lange, W. (1988) Cereal cytogenetics in retrospect. What came true of some cereal cytogenetics' pipe dreams? *Euphytica* **5**, 7-25.
- Langer-Safer, P., Levine, M., and Ward, D. C. (1982) Immunocytological method for mapping genes on *Drosophila* polytene chromosomes. *Proc. Natl. Acad. Sci. USA* **79**, 4381-4385.

- Le, H. T. and Armstrong, K. C. (1991) In situ hybridization as a rapid means to assess meiotic pairing and detection of alien DNA transfers in interphase cells of wide crosses involving wheat and rye. *Mol. Gen. Genet.* **225**, 33-37.
- Lehler, H., Busch, W., Martin, R., and Herrmann, R. G. (1993) Localization of the B-hordein locus on barley chromosomes using fluorescence in situ hybridization. *Chromosoma* **102**, 428-432.
- Leitch, A. R., Schwarzacher, T., and Leitch, I. J. (1994) The use of fluorochromes in the cytogenetics of the small-grained cereales (Triticeae). *Histochemical Journal* **26**, 471-479.
- Leitch, A. R., Schwarzacher, T., Mosgoller, W., Bennett, M. D., and Heslop-Harrison, J. S. (1991) Parental genome are separated throughout the cell cycle in plant hybrid. *Chromosoma* **101**, 206-213.
- Leitch I. J. and Bennett M.D. (1997) Polyploidy in angiosperms. *Trends in Plant Science* **2**, 470-476
- Leitch, I. J. and Heslop-Harrison, J. S. (1992) Physical mapping of the 18S-5.8S-26S rRNA genes in barley by *in situ* hybridization. *Genome* **35**, 1013-1018.
- Leitch, I. J. and Heslop-Harrison, J. S. (1993) Physical mapping for four sites of 5S rDNA sequences and one side of the alfa-amylase gene in barley (*Hordeum vulgare*). *Genome* **36**, 517-523.
- Leitch, I. J., Leitch, A. R., and Heslop-Harrison, J. S. (1991a) Physical mapping of plant DNA sequences by simultaneous *in situ* hybridization of two differently labelled fluorescent probes. *Genome* **34**, 329-333.
- Leitch, I. J., Leitch, A. R., Schwarzacher, T., Maluszynska, J., Anamthawat-Jonsson, K., Shi, M., Harrison, G., and Heslop-Harrison, J. S. (1991b) Two-colour mapping of plant DNA sequences using digoxigenin and biotin. *Boehringer Mannheim Update* **4**, 10-11.
- Lichter, P. (1997) Multicolor FISHing: what's the catch? *Trends in Genet.* **13**, 475-479.
- Lim, K.Y., Leitch I.J., Leitch, A.R. (1998) Genomic characterisation and the detection of raspberry chromatin in polyploid *Rubus*. *Theor. Appl. Genet.* **97**, 1027-1033.
- Lim, K.Y., Matyasek, R., Lichtenstein, C.P. and Leitch A.R. (2000) Molecular cytogenetic analyses and phylogenetic studies in the *Nicotiana* section Tomentosae. *Chromosoma* **109**, 245-258
- Lima-Brito, J., Guedes-Pinto, H., Harrison, G. E., and Heslop-Harrison, J. S. (1997) Molecular cytogenetic analysis of durum wheat x tritordeum hybrids. *Genome* **40**, 362-369.
- Lin, X., Kaul, S., Rounsley, S., Shea, T. P., Benito, M. I., Town, C. D., Fujii, C. Y., Mason, T., Bowman, C. L., Barnstead, M., Feldblyum, T. V., Buell, C. R., Ketchum, K. A., Lee, J., Ronning, C. M., Koo, H. L., Moffat, K. L., Cronin, L. A., Shen, M., Pal, G., Van Aken, S., Umayam, L., Tallon, L. J., Gill, J. E., Adams, M. D., Carrera, A. J., Creasy, T. H., Goodman, H. M., Somerville, C. R., Copenhaver, G. P., Preuss, D., Niernan, W. C., White, O., Eisen, J. A., Salzberg, S. L., Fraser, C. M., and Venter, J. C. (1999) Sequence and analysis of chromosome 2 of the plant *Arabidopsis thaliana*. *Nature* **402**, 761-768.
- Linde-Laursen, I. and Bothmer, R. (1988) Elimination and duplication of particular *Hordeum vulgare* chromosomes in aneuploid interspecific *Hordeum* hybrids. *Theor. Appl. Genet.* **76**, 897-908.
- Lukaszewski, A. J. and Gustafson, J. P. (1983) Translocations and modifications of chromosomes in Triticale x wheat hybrides. *Theor. Appl. Genet.* **64**, 239-248.
- Maluszynska, J. and Heslop-Harrison, J. S. (1991). Localization of tandemly repeated DNA sequences in *Arabidopsis thaliana*. *Plant J.* **1**, 159-166.
- Maluszynska, J. and Schweizer, D. (1989) Ribosomal RNA genes in B chromosomes of *Crepis capillaris* detected by non-radioactive *in situ* hybridization. *Heredity* **62**, 59-65.

- Martin, A., Rubiales, D., and Cabrera, A. (1998) Meiotic pairing in a trigeneric hybrid *Triticum tauschii*-*Agropyron cristatum*-*Hordeum chilense*. *Heredity* **129**, 113-118.
- Martinez-Zapater, J. M., Estelle, M. A., and Somerville, C. R. (1986) A highly repeated DNA sequence in *Arabidopsis thaliana*. *Mol. Gen. Genet.* **204**, 417-423.
- Martini, G. and Flavell, R. (1985) The control of nucleolus volume in wheat, a genetic study at three developmental stages. *Heredity* **54**, 111-120.
- Mesbah, M., Wennekes-van Eden, J., Hans de Jong, J., De Bock, T. S. M., and Lange, W. (2000) FISH to mitotic chromosomes and extended DNA fibres of *Beta procumbens* in a series of monosomic additions to beet (*B. vulgaris*). *Chromosome Res.* **8**, 285-293.
- Miller, J. T., Dong, F., Jackson, S. A., Song, J., and Jiang, J. (1998) Retrotransposon-related DNA sequences in the centromeres of grass chromosomes. *Genetics* **150**, 1615-1625.
- Mikhailova E.I., Naranjo T., Shepherd K., Wennekes-van Eden, J., Heyting C., de Jong J.H. (1998) The effect of the wheat Ph1 locus on chromatin organisation and meiotic chromosome pairing analysed by genome painting. *Chromosoma* **107**: 339-350.
- Molnar-Lang, M., Linc, G., Friebe, B. R., and Sutka, J. (2000) Detection of wheat-barley translocations by genomic *in situ* hybridization in derivatives of hybrids multiplied *in vitro*. *Euphytica* **112**, 117-123.
- Moore, G., Lucas, H., Batty, N., and Flavell, R. (1991) A family of retrotransposons and associated genomic variation in wheat. *Genomics* **10**, 461-468.
- Morais-Cecilio, L., Delgado, M., Jones R.N., Viegas, W. (2000) Modification of wheat rDNA loci by B chromosomes: a chromatin organization model. *Chromosome Research* **8**: 341-371.
- Moscone, E., Klein, F., Lambrou, M., Fuchs, J., and Schweizer, D. (1999) Quantitative karyotyping and dual-color FISH mapping of 5S and 18S-25S rDNA probes in the cultivated *Phaseolus species* (Leguminosae). *Genome* **42**, 1224-1233.
- Moscone, E., Matzke, M. A., and Matzke, A. J. M. (1996) The use of combined FISH/GISH in conjunction with DAPI counterstaining to identify chromosomes containing transgene inserts in amphidiploid tobacco. *Chromosoma* **105**, 231-236.
- Mukai, Y., Endo, T. R., and Gill, B. S. (1991) Physical mapping of the 18S.26S rRNA multigene family in common wheat: Identification of a new locus. *Chromosoma* **100**, 71-78.
- Mukai, Y., Friebe, B., Hatchett, J. H., Yamamoto, M., and Gill, B. S. (1993a) Molecular cytogenetic analysis of radiation-induced wheat-rye terminal and intercalary chromosomal translocations and the detection of rye chromatin specifying resistance to Hessian fly. *Chromosoma* **102**, 88-95.
- Mukai, Y., Nakahara, Y., and Yamamoto, M. (1993b) Simultaneous discrimination of the three genomes in hexaploid wheat by multicolor fluorescence *in situ* hybridization using total genomic and highly repeated DNA probes. *Genome* **36**, 489-494.
- Murata, M., Heslop-Harrison, J. S., and Motoyoshi, F. (1997) Physical mapping of the 5S ribosomal RNA genes in *Arabidopsis thaliana* by multi-color fluorescence *in situ* hybridization with cosmid clones. *Plant J.* **12**, 31-37.
- Ohmido, N., Akryama, Y., and Fukui, K. (1998) Physical mapping of unique nucleotide sequences on identified rice chromosomes. *Plant Mol.Biol.* **38**, 1043-1052.
- Ohmido, N. and Fukui, K. (1997) Visual verification of close disposition between a rice A genome-specific DNA sequence (TrsA) and the telomere sequence. *Plant Mol. Biol.* **35**, 963-968.
- Ohmido, N., Kijima, K., Akiyama, Y., de Jong, J. H., and Fukui, K. (2000) Quantification of total genomic DNA and selected repetitive sequences reveals concurrent changes in different DNA families in *indica* and *japonica* rice. *Mol. Gen. Genet.* **263**, 388-394.
- Orgaard, M. and Heslop-Harrison, J. S. (1994) Investigation of genome relationships between *Leymus*, *Psathyrotachys* and *Hordeum* by genomic DNA:DNA *in situ* hybridization. *Ann. Bot.* **73**, 195-203.

- Orgaard, M., Jacobsen, N., and Heslop-Harrison, J. S. (1995) The hybrid origin of two cultivars of *Crocus* (Iridaceae) analysed by molecular cytogenetics including genomic Southern and *in situ* hybridization. *Ann. Bot.* **76**, 253-262.
- Osuji, J. O., Harrison, G., Crouch, J., and Heslop-Harrison, J. S. (1997) Identification of the genomic constitution of *Musa* L. (bananas, plantains and hybrids) using molecular cytogenetics. *Ann. Bot.* **80**, 787-793.
- Pardue, M. L. and Gall, J. G. (1969) Molecular hybridization of radioactive DNA to DNA of cytological preparation. *Proc. Natl. Acad. Sci. USA* **64**, 600-604.
- Pearce, S. R., Harrison, G., Li, D., Heslop-Harrison, J. S., Kumar, A., and Flavell, A. J. (1996a) The *Ty1-copia* group retrotransposons in *Vicia* species: copy number, sequence heterogeneity and chromosomal localization. *Mol. Gen. Genet.* **250**, 305-315.
- Pearce, S. R., Pich, U., Harrison, G., Flavell, A. J., Heslop-Harrison, J. S., Schubert, I., and Kumar, A. (1996b) The *Ty1-copia* group retrotransposons of *Allium cepa* are distributed throughout the chromosomes but are enriched in the terminal heterochromatin. *Chromosome Res.* **4**, 365-371.
- Pearce, S. R., Harrison, G., Heslop-Harrison, J. S., Flavell, A. J., and Kumar, A. (1997) Characterization and genomic organization of *Ty1-copia* group retrotransposons in rye (*Secale cereale*). *Genome* **40**, 617-625.
- Pedersen, C., Zimny, J., Becker, D., Jähne-Gärtner, A., and Lörz, H. (1997) Localization of introduced genes on the chromosomes of transgenic barley, wheat and triticale by fluorescence *in situ* hybridization. *Theor. Appl. Genet.* **94**, 749-757.
- Pedersen, C., Giese, H., and Linde-Laursen, I. (1995) Towards an integration of the physical and the genetic chromosome maps of barley by *in situ* hybridization. *Hereditas* **123**, 77-88.
- Pickering, R. A., Malyshev, S., Kunzel, G., Johnston, P. A., Korzun, V., Menke, M., and Schubert, I. (2000) Locating introgressions of *Hordeum bulbosum* chromatin within the *H. vulgare* genome. *Theor. Appl. Genet.* **100**, 27-31.
- Pinkel, D., Straume, T., and Gray, J. W. (1986) Cytogenetic analysis using quantitative, high-sensitivity, fluorescence hybridization. *Proc. Natl. Acad. Sci. USA* **83**, 2934-2938.
- Poggio, L., Confalonieri, V., Comas, C., Cuadrado, A., Jouve, N., and Naranjo, C. A. (1999) Genomic *in situ* hybridization (GISH) of *Tripsacum dactyloides* and *Zea mays* ssp. *mays* with B chromosomes. *Genome* **42**, 687-691.
- Presting, G. G., Malysheva, L., Fuchs, J., and Schubert, I. (1998) A *TY3/GYPSY* retrotransposon-like sequence localizes to the centromeric regions of cereal chromosomes. *Plant J.* **16**, 721-728.
- Puzina, J., Javornik, B., Bohanec, B., Schweizer, D., Maluszynska, J., and Papes, D. (1999) Random amplified polymorphic DNA analysis, genome size, and genomic *in situ* hybridization of triploid viviparous onions. *Genome* **42**, 1208-1216.
- Rayburn, A. L. and Gill, B. S. (1985) Use of biotin-labeled probes to map specific sequences on wheat chromosomes. *J. Hered.* **76**, 78-81.
- Ribeiro-Carvalho, C., Guedes-Pinto, H., Harrison, G., and Heslop-Harrison, J. S. (1997) Wheat-rye chromosome translocations involving small terminal and intercalary rye chromosome segments in the Portuguese wheat landrace Barbela. *Heredity* **78**, 539-546.
- Richards, E. J. and Ausubel, F. M. (1988) Isolation of higher eukaryotic telomere from *Arabidopsis thaliana*. *Cell* **53**, 127-136.
- Richards, E. J., Goodman, H. M., and Ausubel, F. M. (1991) The centromere region of *Arabidopsis thaliana* chromosome I contains telomere-similar sequences. *Nucleic Acids Res.* **19**, 3351-3357.
- Schmidt, T. and Heslop-Harrison, J. S. (1996) The physical and genomic organization of microsatellites in sugar beet. *Proc. Natl. Acad. Sci. USA* **93**, 8761-8765.

- Schmidt, T. and Heslop-Harrison, J. S. (1996) The physical and genomic organization of microsatellites in sugar beet. *Proc. Natl. Acad. Sci. USA* **93**, 8761-8765.
- Schmidt, T., Kubis, S., and Heslop-Harrison, J. S. (1995) Analysis and chromosomal localization of retrotransposon in sugar beet (*Beet vulgaris* L.): LINEs and *Ty1*-copialike elements as major components of the genome. *Chromosome Res.* **3**, 335-345.
- Schmidt, T., Kubis, S., Katsiotis, A., Jung, C., and Heslop-Harrison, J. S. (1998) Molecular and chromosomal organization of two repetitive DNA sequences with intercalary locations in sugar beet and other *Beta* species. *Theor. Appl. Genet.* **97**, 696-704.
- Schubert I. and Wobus U. (1985) In situ hybridization confirms jumping nucleolus organizing regions in *Allium*. *Chromosoma* **92**:143-148
- Schwarzacher, H. G., Mikelsaar, A. V., and Schnedl, W. (1978) The nature of the Ag-staining of nucleolus organizer regions: electron- and light-microscopic studies on human cells in interphase, mitosis and meiosis. *Cytogenet. Cell Genet.* **20**, 24-39.
- Schwarzacher, T., Ananthawat-Jonsson, K., Harrison, G. E., Islam, A. K. M. R., Jia, J. Z., King, I. P., Leitch, A. R., Miller, T. E., Reader, S. M., Rogers, W. J., Shi, M., and Heslop-Harrison, J. S. (1992) Genomic *in situ* hybridization to identify alien chromosomes and chromosome segments in wheat. *Theor. Appl. Genet.* **84**, 778-786.
- Schwarzacher, T. and Heslop-Harrison, J. S. (1991) *In situ* hybridization to plant telomeres using synthetic oligomers. *Genome* **34**, 317-323.
- Schwarzacher, T. and Heslop-Harrison, J. S. (2000) *Practical in situ Hybridization*, BIOS, Oxford.
- Sharma A.K. and Sharma A. (1999) Plant Chromosomes: Analysis, Manipulation and Engineering. Harwood Academic Publishers, Amsterdam.
- Shishido, R., Sano, Y., and Fukui, K. (2000) Ribosomal DNAs: an exception to the conservation of gene order in rice genomes. *Mol.Gen.Genet.* **263**, 586-591.
- Soltis, P. S. and Soltis, D. E. (2000) The role of genetic and genomic attributes in the success of polyploids. *Proc. Natl. Acad. Sci. USA* **97**, 7051-7057.
- Suoniemi, A., Ananthawat-Jonsson, K., Arna, T., and Schulman, A. H. (1996) Retrotransposon *BARE-1* is a major, dispersal component of the barley (*Hordeum vulgare* L.) genome. *Plant Mol. Biol.* **30**, 1321-1329.
- Svitashev, S., Ananiev, E., Pawlowski, W. P., and Somers, D. A. (2000) Association of transgene integration sites with chromosome rearrangements in hexaploid oat. *Theor. Appl. Genet.* **100**, 872-880.
- Takahashi, C., Leitch, I. J., Ryan, A., Bennett, M. D., and Brandham, P. E. (1997) The use of genomic *in situ* hybridization (GISH) to show transmission of recombinant chromosomes by a partially fertile bigeneric hybrid, *Gasteria lutzii* x *Aloe aristata* (Aloaceae), to its progeny. *Chromosoma* **105**, 342-348.
- Takahashi, C. Marshall, J.A., Bennett, M.D., Leitch I.J. (1999) Genomic relationships between maize and its wild relatives. *Genome* **42**: 1201-1207.
- Taketa, S., Ando, H., Takeda, K., and Harrison, G. E. (2000) The distribution, organization and evolution of two abundant and widespread repetitive DNA sequences in the genus *Hordeum*. *Theor. Appl. Genet.* **100**, 169-176.
- Tang, S., Li, Z., Jia, X., and Larkin, P. J. (2000) Genomic *in situ* hybridization (GISH) analyses of *Thinopyrum intermedium*, its partial amphiploid Zhong 5, and disease-resistant derivatives in wheat. *Theor. Appl. Genet.* **100**, 344-352.

- ten Hoopen, R., Montijn, B. M., Veuskens, J., Oud, J. L., and Nanninga, N. (1999) The spatial localization of T-DNA insertions in petunia interphase nuclei: consequences for chromosome organization and transgene insertion sites. *Chromosome Res.* **7**, 611-623.
- ten Hoopen, R., Robbins, T. P., Fransz, P. F., Montijn, B. M., Oud, O., Gerats, A. G. M., and Nanninga, N. (1996) Localization of T-DNA insertions in petunia by fluorescence *in situ* hybridization: physical evidence for suppression of recombination. *Plant Cell* **8**, 823-830.
- Thomas, H. M., Morgan, W. G., Meredith, M. R., Humphreys, M. W., Thomas, H., and Leggett, J. M. (1994) Identification of parental and recombined chromosomes in hybrid derivatives of *Lolium multiflorum* x *Festuca pratensis* by genomic *in situ* hybridization. *Theor. Appl. Genet.* **88**, 909-913.
- Vershinin, A.V., Alkhimova, E.G. and Heslop-Harrison, J.S. (1996) Molecular diversification of tandemly organized DNA sequences and heterochromatic chromosome regions in some Triticeae species. *Chrom. Res.* **4**: 517-525.
- Vershinin, A.V., Schwarzacher, T. and Heslop-Harrison, J.S. (1995) The large-scale genomic organization of repetitive DNA families at the telomeres of rye chromosomes. *Plant Cell* **7**, 1823-1833
- Wang, Z. X., Kurata, N., Katayose, Y., and Minobe, Y. (1995) A chromosome 5-specific repetitive DNA sequence in rice (*Oryza sativa* L.). *Theor. Appl. Genet.* **90**, 907-913.
- Weiss H. and Maluszynska J. (2000) Chromosomal rearrangement in autotetraploid plants of *Arabidopsis thaliana*. *Hereditas* **133**:255-261
- Weiss, H., Pasierbek, P., and Maluszynska, J. (1999). An improved nonfluorescent detection system for *in situ* hybridization in plants. *Biotechnic and Histochemistry* **75**, 49-53.
- Wolters, A.-M. A., Trindade, L. M., Jacobsen, E., and Visser, R. G. F. (1998) Fluorescence *in situ* hybridization on extended DNA fibres as a tool to analyse complex T-DNA loci in potato. *Plant J.* **13**, 837-847.
- Xu, J. and Earle, E. D. (1994) Direct and sensitive fluorescence *in situ* hybridization of 45S rDNA on tomato chromosomes. *Genome* **37**, 1062-1065.
- Yang, Q. Hanson, L. Bennett, M.D. and Leitch, I.J. (1998) Genome structure and evolution in the allohexaploid weed *Avena fatua* L. (Poaceae). *Genome* **42**: 512-518.
- Zhong, X. B., Fransz, P. F., Wennekes-van Eden, J., Ramanna, M. S., Van Kammen, A., Zabel, P., and de Jong, J. H. (1998) FISH studies reveal the molecular and chromosomal organization of individual telomere domains in tomato. *Plant J.* **13**, 507-517.
- Zhong, X. B., Hans de Jong, J., and Zabel, P. (1996) Preparation of tomato meiotic pachyten and mitotic metaphase chromosomes suitable for fluorescence *in situ* hybridization (FISH). *Chromosome Res.* **4**, 24-28.

MOLECULAR TOOLS FOR IMPROVING COFFEE (*COFFEA ARABICA* L.) RESISTANCE TO PARASITES

DIANA FERNANDEZ and PHILIPPE LASHERMES

IRD (Institut de Recherche pour le Développement)

BP 5045, 34032 Montpellier cedex, FRANCE

1. INTRODUCTION

Coffee is so far the world's most valuable agricultural export commodity, contributing over 10-11 billion US dollars annually (Bolvenkel *et al.* 1993). In particular, coffee represents one the key export and cash crops in tropical and subtropical developing countries with generally a favourable impact on the social and physical environment.

Although the genus *Coffea* is reported to comprise over 80 species (Bridson and Verdcourt, 1988), only two species *Coffea arabica* L. popularly called as arabica and *C. canephora* Pierre, known as robusta are under commercial cultivation. Arabica, the high land coffee accounts for nearly 70 per cent of global production while robusta coffee is more adaptable for low lands and contributes remaining 30 per cent. Another species *Coffea liberica* (liberica coffee) stands third with a share of less than one per cent of world coffee production. Arabica produces superior quality coffee but the production levels are often constrained by major diseases and pests like leaf rust (*Hemileia vastatrix* Berk & Br.), coffee berry disease (*Colletotrichum kahawae*), coffee berry borer (*Hypothenemus hampei*), stem borer (*Xylotrechus quadripes* Chev.) and nematodes (*Meloidogyne spp.* and *Pratylenchus spp.*). Robusta is more tolerant to these diseases and pests but product quality is rather inferior. Hence, transfer of desirable genes in particular for disease resistance from diploid species like *C. canephora* and *C. liberica* into tetraploid arabica cultivars without affecting quality traits has been the main objective of arabica breeding.

Molecular techniques are valuable tools for improving the efficiency of conventional plant breeding by allowing indirect selection for a trait of interest by looking for molecular markers linked to this particular trait. Such strategy is especially useful for agronomic traits such as resistance to parasites when pathogenicity tests are time-consuming, often destructive for the progenies to evaluate, and difficult to interpret. Availability of tightly linked genetic markers for resistance genes are of great help in selecting plants carrying these genes without subjecting them to pathogen attacks.

In addition, molecular tools offer great scope for assessing the genetic diversity of wild and cultivated species and evaluating plant genetic resource stocks. These analyses might also be extended to parasites species affecting a particular crop in order to gain knowledge on the genetic organization of their natural populations and determine adapted control strategies.

Finally, with the use of molecular techniques, it is now possible to facilitate the transfer of desirable genes among plant varieties. In addition, data derived from the sequencing of model plant genomes provide a growing body of knowledge that is expected to bring up the research on coffee more pertinent and efficient.

In this chapter, we highlight the usefulness of molecular tools in research programs aiming at improving the natural resistance of coffee plants to parasites. We review research aspects linked to the genetic characterization of coffee pathogens, the evaluation of coffee genetic resources, breeding for resistance, as well as cloning and expression of resistance genes in coffee plants.

2. MOLECULAR MARKERS FOR ASSESSING PATHOGENIC POPULATIONS DIVERSITY AND FOR PATHOGEN IDENTIFICATION

The two main strategies that are currently being used for achieving control of pests and parasites are (i) chemical treatments through the use of pesticides during cropping practices, and (ii) deployment of resistant varieties through plant breeding programs. With both forms of disease control, assessment of the level of genetic variability in pathogenic populations is extremely important for disease management and epidemiology (Milgroom and Fry 1997).

Many examples have shown that, under selection pressure, plant parasites may evolve to develop resistance to chemicals or to overcome host resistance genes. The intrinsic level of genetic diversity of the pathogenic populations is a factor that may contribute to the potentiality of parasites to evolve new specificities and to overcome plant resistance genes.

To this regard, knowledge of the pathogenic and genetic diversity of natural populations is indispensable for the efficient management of disease resistance genes (Brown, 1995). Local deployment of appropriate resistance genes in each geographical areas where the disease occurs may lead to an effective and durable control of pathogen populations only if previous characterization of these populations has been conducted. In addition, knowledge of the genetic variation of the pathogen is required to progress in the selection of resistance sources.

The use of DNA molecular markers have made possible more precise identification of the genetic entities composing natural populations (Brown 1996; Leung *et al.* 1993; McDonald 1997). Within parasites there is a great diversity of biological behaviors and life histories that have direct implications for the genetic structure of their populations. Many microorganisms are capable of reproducing both sexually and asexually and the relative contribution of each of these modes of reproduction can have major implications for their genetic population structure (Brygoo *et al.* 1998).

To this regard, diagnostic tests able to identify and to early detect a pathogenic species or a pathogenic group within a species would be of great help. Molecular studies for parasites identification in coffee production are still in their infancy. However, many research programs are currently being developed to integrate new DNA-based technologies for improving coffee pathogen control and it is expected to develop much more. We will review in this chapter some of the recent works that came to our knowledge.

2.1. NEMATODES

Among root-knot nematode species that have been reported to infect coffee plants, the most damaging species are *Meloidogyne exigua*, *M. arabicida*, *M. paranaensis* and *M. incognita* (Campos *et al.*, 1990 ; Carneiro *et al.*, 1996a ; Lopez and Salazar, 1989).

Identification of species within *Meloidogyne* genus was initially based on morphological characteristics and host preferences, but these features are prone to variation and difficult to observe. In addition, because most *Meloidogyne* species reproduce by parthenogenesis, they are not amenable to crossing experiments for inferring genetic relationships between groups. Markers able to differentiate between species, populations and pathotypes are thus requested to help in epidemiological surveys.

Isozyme analysis was found to be the more reliable method to classify *Meloidogyne* species (Esbenshade and Triantaphyllou, 1985), but does not detect much intraspecific variation among nematodes populations (Carneiro *et al.*, 1996b ; Hernandez, 1997). In contrast, molecular markers such as RAPDs and AFLPs have shown potential to separate *Meloidogyne* species based on their DNA amplification patterns and allowed to detect variation between populations (Cenis, 1993 ; Castagnone-Sereno *et al.*, 1994 ; Semblat *et al.*, 1998). Application of these techniques to coffee nematode's populations will help in assessing the frequency and distribution of the different virulent groups and *Meloidogyne* species in coffee growing areas. In addition, once carefully genetically characterized, nematode populations used in pathogenicity tests will provide useful data for coffee resistant sources evaluation.

2.2. COFFEE RUST

A major fungal disease of *C. arabica* plants is the orange rust caused by the basidiomycetous fungus *Hemileia vastatrix* Berkeley and Broome. In nature, this rust fungus only displays a shortened biological cycle with production of uredinial, telial and basidial stages on coffee. No alternate host for the pycnial and aecial stages are known. This biotrophic fungus is thought to only propagates via urediniospores (asexual stage).

Nowadays, at least 40 races of the fungus have been described by artificial inoculation on a set of several coffee species and interspecific hybrids (Kushalappa and Eskes, 1989 ; Rodrigues, 1975 ; Rodrigues, pers. comm.). Some races like race II have a world-wide distribution but others are restricted to specific areas such as in India where a number of indigenous races have been detected. This may be explained by the presence in this country of particular coffee genotypes whereas most arabica cultivars prevalent throughout the world display a narrow genetic basis (Rodrigues, 1975). Because of the obligatory nature of the growth cycle of the fungus, and the large coffee differential set to be artificially inoculated, the identification of rust races still remains a long-lasting process.

The use of molecular tools for achieving rapid characterization of isolates has been tested with RAPDs (Holguin Melendez, 1993 ; Nandris *et al.*, 1998) and PCR-RFLP of the ribosomal ITS (Kohler F., pers. com.). Preliminary results showed that, in

contrast with their high diversity of pathogenic behaviors, the rust isolates exhibited a low level of genetic polymorphism and no direct correlation between polymorphic profiles and pathotypes could be evidenced. The absence of genetic diversity among isolates, as measured using DNA neutral markers, may be interpreted as the result of the asexual propagation of the fungus throughout the world. However, more studies are needed on an extended set of isolates and with molecular markers known to reveal genetic diversity such as microsatellites or AFLPs.

2.3. COFFEE BERRY DISEASE (CBD)

CBD caused by the fungus *Colletotrichum kahawae* is the most important disease of the *C. arabica* crop in various parts of the African continent. CBD is an anthracnose of green and ripe coffee berries, and although *C. kahawae* is morphologically similar to *C. gloeosporioides*, these pathogenic characteristics together with some cultural and biochemical traits distinguish *C. kahawae* from other *C. gloeosporioides*-infecting coffee (Waller *et al.*, 1993).

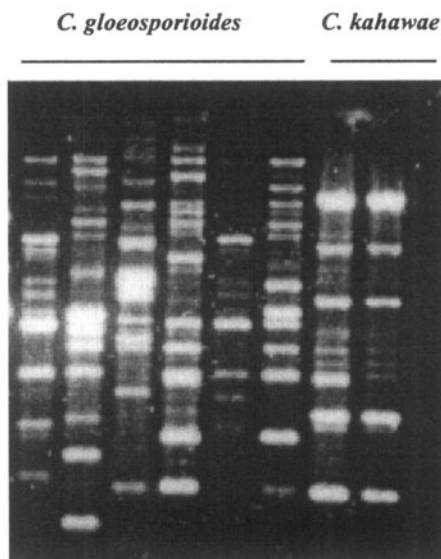


Figure 1. RAPD patterns generated on *Colletotrichum gloeosporioides* and *C. kahawae* isolated from coffee plants (M. Lourd and D. Fernandez, pers. comm.).

Molecular studies confirmed the close genetic relationship of *C. kahawae* to the group species *C. gloeosporioides* but DNA sequence polymorphisms were detected between the two species based on RFLPs of mitochondrial and ribosomal DNAs (Sreenivasaprasad *et al.*, 1993) or RAPDs (M. Lourd and D. Fernandez, pers. comm.) (Figure 1).

Among *C. kahawae* isolates collected in widely distributed sources in Africa, a very low level of genetic variability was observed using either RAPDs (Manga, 1999) or vegetative compatibility groups (Beynon *et al.*, 1995 ; Manga, 1999). However, this analysis revealed a geographic substructuring of isolates, indicating that *C. kahawae* populations in West and East Africa may have evolved separately (Manga, 1999). Variations in pathogenicity of *C. kahawae* isolates have been recorded (Beynon *et al.*,

1995) that, together with the genetic differentiation of populations may have implications for coffee breeding for resistance to CBD.

3. Evaluation of coffee genetic resources

3.1. PHYLOGENETIC RELATIONSHIPS

Coffee-trees are classified in two genera, *Coffea* and *Psilanthus*, each genus being divided into two subgenera (Bridson and Verdcourt, 1988). All *Coffea* species are native to the inter-tropical forest of Africa, Madagascar and islands of the Indian Ocean (Mascareign and Comoro Islands), while species belonging to the genus *Psilanthus* originate from either Asia or Africa.

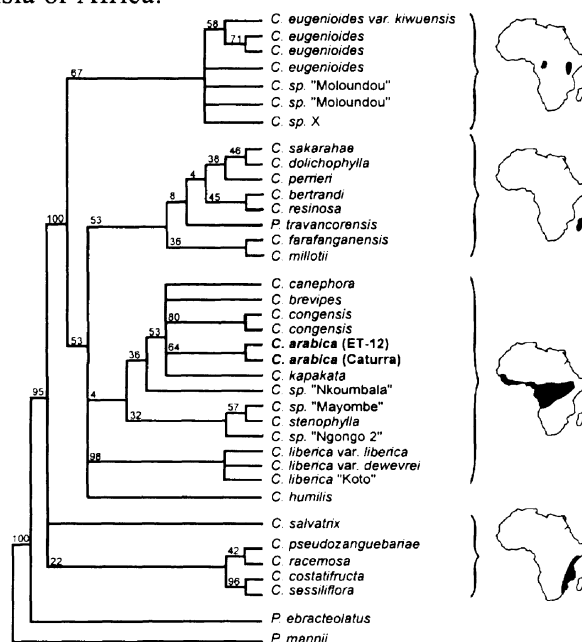


Figure 2. Molecular phylogeny of *Coffea* species. Strict consensus Wagner tree resulting from parsimony analysis of ITS 2 (nuclear ribosomal DNA) sequence data. Numbers above the nodes represent bootstrap values (%). Geographical distributions of the major groups are also indicated (Lashermes et al 1997).

The subgenus *Coffea* encompasses more than 80 taxa so far identified, including the two species of economic importance: *Coffea arabica* L. and *Coffea canephora* Pierre (Charrier and Berthaud, 1985). *Coffea* species are diploid ($2n=2x=22$), except *C. arabica* ($2n=4x=44$) which is self-fertile and allotetraploid (Carvalho, 1952). Molecular phylogeny of *Coffea* species has been established based on DNA sequence data. The internal transcribed spacer (ITS 2) region of the nuclear ribosomal DNA (Lashermes *et al.*, 1997) as well as the chloroplast DNA variation (Cros *et al.*, 1998) were successfully used to infer phylogenetic relationships of *Coffea* species (Figure 2). The results suggest a radial mode of speciation and a recent origin in Africa for the genus *Coffea*. Several major clades were identified, which present a strong geographical correspondence (i.e. West Africa, Central Africa, East Africa and Madagascar). Molecular cytogenetic analyses (Raina *et al.*, 1998 ; Lashermes *et al.*,

1999) have indicated that *C. arabica* is an amphidiploid formed from the hybridisation between two closely related diploid species (i.e. *C. canephora* and *C. eugenoides*). The evidence suggests recent speciation and low divergence between the two constitutive genomes of *C. arabica* and those of its progenitor species. Furthermore, recent investigations suggest that homoeologous chromosomes do not pair in *C. arabica*, not as a consequence of structural differentiation, but because of the functioning of pairing regulating factors (Lashermes *et al.*, 2000a).

3.2. GENOME SIZE

The nuclear DNA content of several coffee species has been estimated by flow cytometry (Cros *et al.*, 1995). The DNA amount (2C values) varies between diploid coffee species from 0.95 to 1.8 pg. In comparison to other angiosperms (Bennett and Leich, 1995), the genomes of coffee species appear to be of rather low size (i.e. 800 and 1200 Mb for *C. canephora* and *C. arabica*, respectively). These variations in DNA amount, other than variation due to ploidy level (e.g. *C. arabica*), are probably due almost entirely to variation in the copy number of repeated DNA sequences. Differences may correspond to genomic evolution correlated with an ecological adaptation process. Furthermore, reduced fertility of certain interspecific F₁ hybrids appears to be associated with significant differences in nuclear content of parental species.

3.3. MOLECULAR GENETIC MARKERS AND DIVERSITY

In coffee, a whole range of different techniques has been used to detect polymorphisms at the DNA level, including randomly amplified polymorphic DNA (RAPD) (Orozco-Castillo *et al.*, 1994 ; Lashermes *et al.*, 1996a ; Rani *et al.*, 2000 ; Anthony *et al.*, 2001), cleaved amplified polymorphisms (CAP) (Lashermes *et al.*, 1996b ; Orozco-Castillo *et al.*, 1996), restriction fragment length polymorphisms (RFLP) (Paillard *et al.*, 1996, Dussert *et al.*, 1999, Lashermes *et al.*, 1999), amplified fragment length polymorphism (AFLP) (Lashermes *et al.*, 2000b) and simple sequence repeats (SSR) or microsatellites (Mettulio *et al.*, 1999 ; Combes *et al.*, 2000). The use of molecular methods has opened up new possibilities for genetic analysis and provides new tools for the efficient conservation and use of coffee genetic resources.

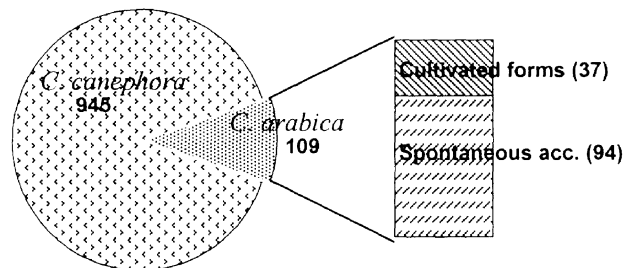


Figure 3. Assessment of the plant genetic diversity. Pie charts depicting the numbers of polymorphic markers (AFLP) observed among individuals within each group constituted by representative accessions of *C. arabica* and *C. canephora*, respectively. For *C. arabica*, cultivated forms (i.e. Bourbon- and Typica-derived cultivars) and wild accessions from Ethiopia were distinguished (Lashermes *et al.* 2000b).

For instance, the genetic diversity in *Coffea arabica* L. appeared extremely reduced in comparison to the diversity observed in *C. canephora* (Figure 3). This low genetic diversity has been attributed to the allotetraploid origin, reproductive biology and evolution process of *C. arabica*. However, spontaneous accessions collected in the primary centre of diversity (i.e. Ethiopian plateau) appeared to constitute a valuable gene reservoir. Indeed, the polymorphism detected among the spontaneous accessions was much higher than among the cultivated accessions.

Furthermore, the relationships between the cultivars and the accessions collected in the primary centre of diversity (Figure 4) as well as the main steps of coffee domestication were clarified based on molecular marker genetic diversity analysis (Anthony *et al.*, 2001). Since heterosis effect has been reported in F₁ hybrid resulting from crosses between subspontaneous Ethiopian accessions and improved cultivars, the possibility of employing marker-based genetic distances for predicting hybrid performance should be considered (Lashermes *et al.*, 1996a ; Bertrand *et al.*, 1999).

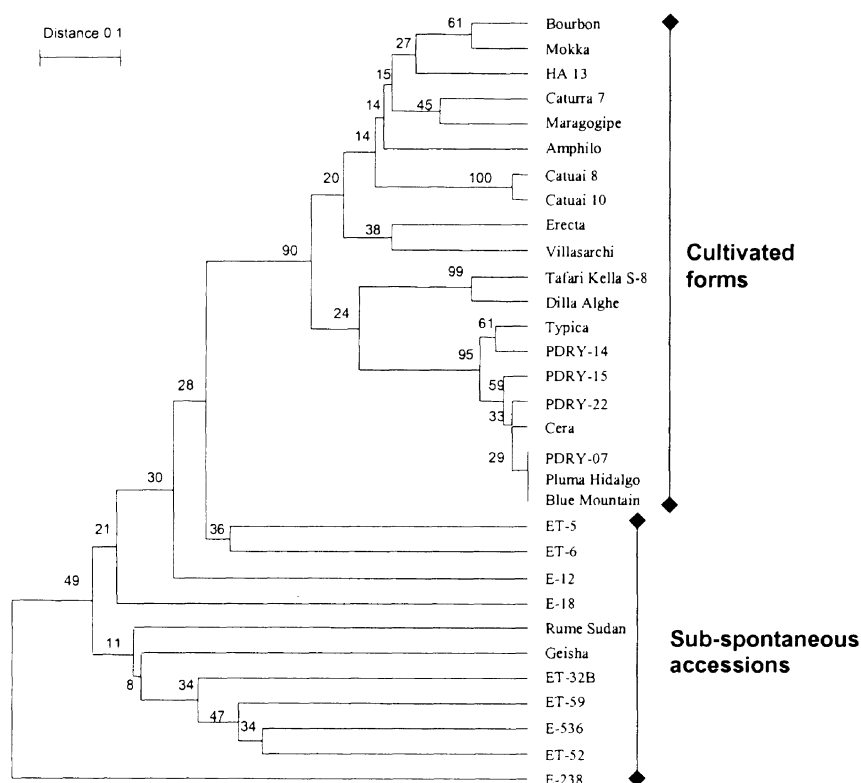


Figure 4. Diversity analysis. Dendrogram of the *C. arabica* accessions generated by group average clustering (UPGMA) using AFLP-based genetic distance. Numbers below branches are bootstrap values (%).

4. Breeding for resistance

Conservation and evaluation of coffee genetic resources has long been considered as an essential part of coffee genetic improvement. Spontaneous accessions of *C. arabica*

collected in the primary centre of diversity as well as wild relative *Coffea* species constitute a valuable gene reservoir for different breeding purposes. For instance, several major resistance genes to leaf rust were identified in *C. arabica* and in progenies derived from interspecific hybrids between *C. arabica* and either *C. canephora* or *C. liberica* (Bettencourt and Rodrigues 1988). Resistance's to Coffee Berry Disease (CBD) and Bacterial Blight of Coffee (BBC, caused by *pseudomonas syringae*) were identified in spontaneous accessions of *C. arabica* while a resistance to leaf miner (*Perileucoptera coffeella*) was identified in *C. racemosa* (Van der Vossen and Walyaro, 1980 ; Guerrero Filho *et al.*, 1999). Furthermore, resistance's to root-knot nematodes were identified in spontaneous accessions of *C. arabica* (i.e. *M. incognita*) as well as in *C. canephora* (i.e. *M. exigua* , *M. paranaensis*) (Anzueto *et al.*, 2001; Gonçalves and Pereira, 1998; Sera *et al.*, 2000).

During the last century, breeding for disease and pest resistance has greatly contributed to the improvement of coffee production. However, conventional coffee breeding methodology faces considerable difficulties when transferring resistance traits from spontaneous accessions or wild diploid species into tetraploid arabica cultivars. The main source of resistance to pests and diseases which has been extensively used world wide, is constituted by the Timor Hybrid, a natural hybrid between *C. arabica* and *C. canephora*. Its exploitation have so far relied on conventional procedures in which a hybrid is produced between an outstanding variety and a donor genotype carrying the trait of interest, and the progeny is backcrossed to the recurrent parent. Undesirable genes from the donor parent are gradually eliminated by selection. In so doing, conventional coffee breeding methodology faces considerable difficulties.

The ploidy difference between *C. arabica* and the donor species, the lack of information regarding genome recombination in interspecific hybrids, the low accuracy of current strategy, the long generation time of coffee-tree, and the high cost of field trial are among the most serious limitations. One can estimate that a minimum of 25 years after hybridisation is required to restore the genetic background of the recipient cultivar and there by ensure good quality of the improved variety. Combining various genes of resistance without reducing coffee quality appears therefore as a very difficult task in an acceptable time-frame through traditional breeding approaches.

4.1. MOLECULAR ASSISTED BACKCROSS BREEDING

Introgressed arabica genotypes derived from the Timor Hybrid were analysed for the presence of *C. canephora* genetic material using AFLP approach (Lashermes *et al.*, 2000b). Although varying between the Timor Hybrid-derived genotypes, the amount of alien genetic material appeared substantial. The introgressed chromosome segments did not appear to be eliminated or counter-selected during the process of selfing and selection. The development of adapted breeding strategies would be particularly suitable. For example, elimination of donor parent chromosomes unlinked to the target gene as well as donor parent segments which are linked to the target gene could be considerably facilitated by the use of molecular markers.

A genome selection could be performed by the use of markers scattered throughout the genome resulting in a reduction of the number of backcross generations required to restore the genetic background of the recipient cultivar. Values were estimated for a theoretical arabica genome of 22 chromosome pairs of, on average, 100

cM each (Total genome of 2200 cM) using equations developed by Hillel *et al.* (1990) and Hospital *et al.* (1992). In the absence of selection, parental donor DNA is only removed by a factor of two in each generation. Simulations are given for MAS programme in which the either 10 or 2% best (in terms of percent recurrent parent genome) individuals in each generation were used as the parent for the next generation (Figure 5). Results equivalent to BC5 generation without selection are obtained after only two marker-assisted BC generations allowing a considerable time saving.

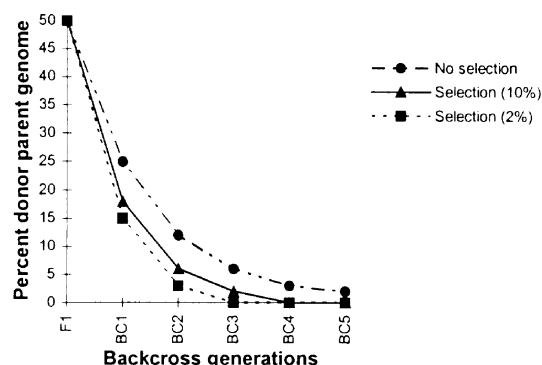


Figure 5. Average content (%) of the donor parental genome in backcross generations under various intensities of genomic selection.

Removing of the linked donor segment could take numerous generations (Stam and Zeven, 1981). Many examples of "linkage drag" are known in which undesirable traits that are closely linked to a target gene are carried out along during breeding programme. For instance in Arabica, even after 6 backcross generations, a region of 32cM flanking a target gene is expected to persist (Lashermes *et al.*, 2000c). In most plant genomes 32cM is enough DNA to contain hundreds of genes. DNA markers could be used to eliminate, or at least significantly reduce, linkage drag by allowing the identification of rare recombinant individuals which are usually only selected by chance in classical breeding.

4.2. SELECTION FOR RESISTANCE GENES THROUGH LINKAGE WITH MOLECULAR MARKERS

An application of molecular marker in coffee breeding is based on finding tight linkages between these markers and target genes of interest. Once identified, it may be much more efficient to select for the marker than for the trait itself. Marker-assisted selection (MAS) of major genes as well quantitative trait loci (QTL) of resistance can be considered. Benefits obtained from marker-selection depend on several factors such as the degree of linkage between the marker and the target gene, savings in time, and the relative costs of direct vs. marker-facilitated selection (Melchinger 1990). There are several reasons for using maker genes to improve selection efficiency in coffee. However, MAS shows undoubtedly considerable interests for the transfer of resistance genes in a variety of circumstances such as:

* *Quarantined pathogens*

If a virulent pathogen does not naturally occur in the test environment, artificial inoculation is prohibited for safety reasons. For instance, CBD is still restricted to the continent of Africa, and the availability of markers linked to the resistance gene(s) could allow pre-emptive breeding in countries (Asia, Latin America) where quarantine barriers are still effective.

* *Reliability/limitation of direct testing for the resistance trait*

Conventional selection progress could be hampered by the difficulty to ensure reliable test. Seedling test could also present strong inconvenient. For instance, the present seedling test of evaluation for root-knot nematode is destructive leading to important difficulties in the utilisation of identified plant resistance sources (Bertrand *et al.* 1995). In addition, expression of many resistance genes can be strongly influenced by environmental conditions.

* *Transfer of recessive resistance genes*

The classical procedure of transferring a recessive resistance gene (i.e. resistance factors to CBD ; Van der Vossen and Walyaro, 1980) includes a progeny test after each backcross generation to determine the presence of the desired allele. With MAS, the transfer can be accomplished without interruptions leading to an important time saving.

* *Earlier and more intense selection*

Another relevant reason to use marker genes is to reduce the time from sowing to selection. Markers may be detectable in seedlings within days of sowing, so avoiding the waste of valuable resources involved in growing coffee trees, most of which may prove to be of no value to the breeding programme. Selection at a juvenile stage may also allow much larger populations to be studied and hence more intense selection to be applied.

* *Pyramiding of resistance genes/Combining valuable traits*

Pyramiding of resistance genes has been suggested as a strategy to provide durable resistance (i.e. coffee leaf rust, CBD, root-knot nematode). However, conventional breeding is complicated by the fact that, it is difficult or often impossible to distinguish the various resistance genotypes. Once the different genes conferring resistance to the same pathogen are tagged by tightly linked marker, they could be relatively easily be accumulated into a single genotype via marker-facilitated selection. Comparable advantages versus conventional are procured when trying to combine simultaneously resistance genes to different disease/pests.

4.3. RESISTANCE GENE MAPPING IN COFFEE

An important step in genetic analysis is to produce linkage genetic maps. Such maps represent the relative order of genetic markers, and their relative genetic distances from one to another, along each chromosome of an organism (Paterson *et al.* 1991). A complete linkage map of *C. canephora* (Figure 6) and an interspecific partial linkage map (*C. pseudozanguebarie* x *C. liberica*) have been so far reported (Paillard *et al.*, 1996 ; Ky *et al.*, 2000 ; Lashermes *et al.*, 2001). Additional linkage maps on different coffee species including *C. arabica* are being constructed. In so doing, the development

of microsatellite PCR-based markers will be particularly useful (Rovelli *et al.*, 2000). These genetic linkage maps bring important information's on coffee genome and chromosomal organisation. In particular, one might use it to map important genes (e.g. the S-locus controlling self-incompatibility ; Lashermes *et al.*, 1996c) as well as loci underlying quantitative traits (i.e. QTL) such as biochemical compounds implied in coffee cup quality (Barre *et al.*, 1998).

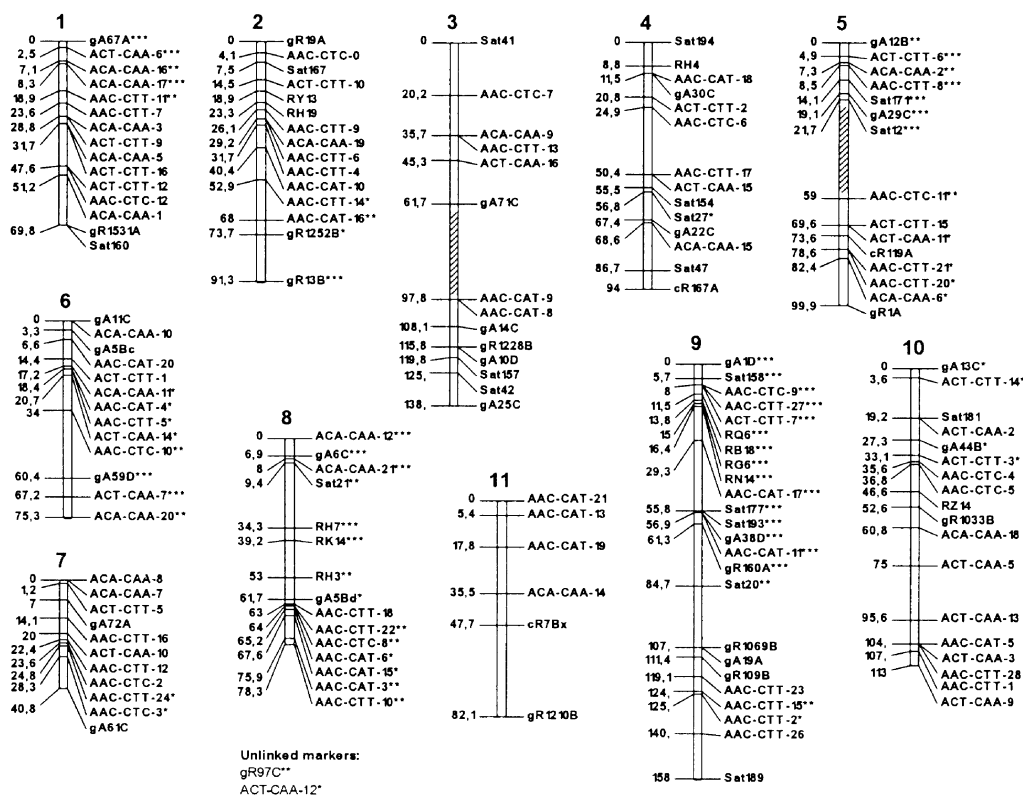


Figure 6. Linkage map of *C. canephora* (Lashermes *et al.*, 2001). Cumulated distances in cM (Kosambi function) are indicated on the left side of the linkage groups. All linkage groups were constituted with a LOD > 5 except two associations indicated by hachures (LOD = 3.5). Loci marked *, **, *** deviated significantly from a 1:1 ratio at $P < 0.05$, $P < 0.01$ and $P < 0.001$, respectively.

Regarding coffee resistance genes, none precise location has been so far reported. However, molecular markers linked to several resistance genes including the *T* resistant gene to CBD (Agwanda *et al.*, 1997), a gene of resistance to *M. exigua* (Figure 7), and resistance factors to leaf rust (Moreno *et al.*, 2000 ; Tedesco *et al.*, 1999), have been already identified.



Figure 7. Example of AFLP Marker linked to the resistance to *M. exigua* derived from the Timor hybrid. Plants showing resistance to *M. exigua* are indicated by R (Resistance).

5. Cloning of resistance genes

Recently, efforts have been directed to the cloning of coffee resistance genes. Comparison of genes for resistance (R) against diverse pathogens from a variety of plants has revealed that many share conserved sequence motifs (Richter and Ronald, 2000). Using degenerate primers based on conserved motifs, nine distinct classes of NBS-like resistance gene analogs (i.e. RGA) representing a large diversity have been already isolated from *Coffea arabica* and *C. canephora* species (Noir *et al.*, 2001). Phylogenetic relationships between deduced amino acid coffee RGA sequences and a representative set of NBS domains of known R-gene products (isolated from *Arabidopsis*, tomato, potato, pepper, lettuce, maize and rice) were investigated (Figure 8). Most coffee RGAs appeared closely related by sequence to at least one known R-gene. The high similarity between particular coffee RGAs and R-genes isolated from other angiosperm species such as *Arabidopsis*, tomato and rice indicated an ancestral relationship and the existence of common ancestors. The maintenance of different R-gene families showing sequence divergence and specific signature might result in fitness superiority and suggests a functional family-specificity. Efforts are being pursued to explore the possibility of implication of isolated coffee RGA families in specific resistance's.

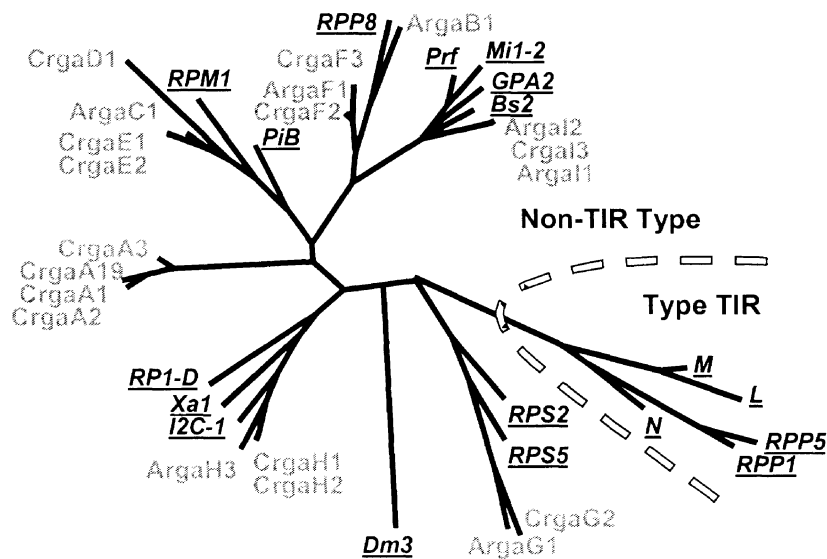


Figure 8. Neighbor-Joining tree based on alignment of amino acid sequences of representative coffee RGAs (1 to 4 members per family) and NBS domains of cloned R-genes (underlined).

Furthermore, positional approach (map-based) for isolation of resistance genes is also investigated. For example, a fine genetic map of the chromosome segment carrying the *M. exigua* resistance gene has been constructed as preliminary step toward positional cloning.

5. Molecular mechanisms of plant disease resistance

The early recognition event between the R protein and the pathogen's avr gene product triggers the rapid activation of several plant defense responses that in concert lead to disease resistance. Among these responses is rapid localized cell death at the site of infection, termed the hypersensitive response (HR), which is thought to limit pathogen invasion in plants.

In the last few years, a systematic search of genes involved in the signal transduction pathways leading to HR and resistance expression has been performed in *Arabidopsis*, tomato and barley. We are now gaining an understanding of the molecular basis of disease resistance specificity and expression in plants. By instance, it has been shown that distinct signal transduction pathways are required by structurally different R proteins. There are also clear evidence for the existence of convergence points in signaling pathways triggered by multiple R genes. Many R genes require additional genes for their function. The Pto pathway in tomato which lead to resistance to *Pseudomonas syringae* is one of the best characterized examples (Sessa and Martin, 2000). Signal transduction originating from the AvrPto-Pto recognition event involves additional proteins, such as Prf and the protein kinase Pti1 which physically interacts with Pto.

Effective expression of resistance would then be jeopardised by transferring a R gene of interest in a genetic background devoid of the appropriate signaling components that should relay the specific recognition information. Although the data generated from *Arabidopsis* and other crops provide insights on the molecular mechanisms that lead to activation of plant defence responses, its direct application for achieving coffee disease control might not be possible.

Recently, an EST (Expressed Sequence Tag) project aiming at characterizing genes early expressed during the hypersensitive reaction of coffee (*C. arabica*) to the rust fungi and nematodes has been initiated (Fernandez *et al.*, 2001). Cloning and functional analysis of genes involved in the *H. vastatrix* and *Meloidogyne* resistance pathways will not only lead to identification of some genetic components controlling disease resistance and cell death, but also allow to manipulate these non-pathogen specific genes to achieve broad spectrum resistance in coffee.

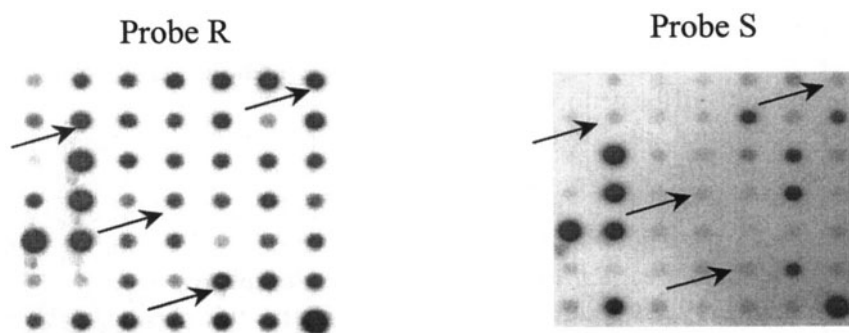


Figure 9. Differential screening of cDNA clones arrayed on Nylon membrane by hybridization against complex cDNA probes generated from resistant (R) and susceptible (S) mRNA pools. Arrows indicate some examples of clones that were selected based on the intense hybridization signal with probe R and the weak signal obtained with probe S.

Using the suppression subtractive hybridization method (Diatchenko *et al.*, 1996), coffee cDNA libraries enriched in sequences specifically expressed during the HR were obtained (Fernandez *et al.*, 2001). In addition, recombinant clones were screened by differential hybridization against complex cDNA probes generated from resistant and susceptible mRNA pools (Figure 9). Sixty per cent of the sequenced clones showed high similarities with plant sequences in databases, and half of them with proteins which role in plant defence reactions have been suggested or demonstrated such as chitinases, heat shock proteins, cytochrome P450, metallothioneins and ionic channels (Choi *et al.* 1996 ; Clough *et al.*, 2000; Guy *et al.*, 1998; Van Loon *et al.* Van Strien, 1999; Whitbred and Schuler, 2000).

Future work will aim at understanding the role of selected clones and sequences showing no similarities with known proteins in the mechanism of coffee resistance to parasites. Genes isolated in coffee plants will be studied for their *in vivo* function and the transcriptional regulation of their expression. Such data may be obtained by transformation of coffee plants with recombinant vector DNA carrying either the gene of interest under control of promoters driving specific gene expression patterns, or the gene promoter fused to a reporter gene that allows accurate quantification and localization of gene expression. However, the whole process of genetic coffee transformation, from primary inoculated explant to plantlet transfer to the greenhouse, usually takes 12 to 20 months (Spiral *et al.*, 1999). In addition, flowering in coffee plants is expected to occur 4 yrs later, only. Such a delay is still an important handicap to consider coffee genetic transformation as a routine tool for functional gene analysis. As an alternative, an heterologous transformation system such as tobacco would be useful for analysis of gene function, but confirmation would still be required in the coffee genome. Transformation of tobacco plants (*Nicotiana tabacum*) with the promoter of the 11S seed storage protein gene (*csp1*) showed it was functional in tobacco and driven the expression of the reporter gene in tobacco grains (Marraccini *et al.*, 1999).

6. Genetic transformation of coffee

Genetic transformation of coffee plants has been successfully achieved by several research groups (Hatanaka *et al.*, 1999 ; Leroy *et al.*, 1999 ; Spiral *et al.*, 1993,1999 ; Sugiyama *et al.*, 1995 ; Van Bortel *et al.*, 1995). Up to now, only a few genes have been transferred into coffee genotypes (Spiral *et al.*, 1999).

Two techniques are available : direct transformation through particule bombardment (biolistics), and *Agrobacterium*-mediated transformation. The first transgenic coffee plants were obtained by co-culturing somatic embryos of *C. arabica*, *C. canephora* and interspecific hybrids with an *A. rhizogenes* vector carrying the reporter gene GUS and the NPTII gene for resistance to kanamycin (Spiral *et al.*, 1993 ; Spiral and Pétiard, 1993). Antibiotic selection (kanamycin) proved to be inefficient (Spiral *et al.*, 1993). In the absence of any other selective marker, the production of hairy roots was used as a control of transformation. Genetic transformation with *A. rhizogenes* was very efficient with a high amount (10 - 40 %) of regenerated roots showing GUS expression (histochemical assay), and confirmation of vector DNA

sequences integration could be generally obtained in regenerated embryos by PCR detection and Southern blotting (Spiral *et al.*, 1999 ; Spiral and Pétiard, 1993). However, when transferred to the greenhouse, disturbed phenotypes (crinkled leaves, short internodes) were observed in regenerated plantlets, due to the expression of various hairy-root genes. *C. canephora* transformed plants developed the more disturbed phenotypes (Spiral *et al.*, 1999). The transformation protocol was further transferred from *A. rhizogenes* to a disarmed *A. tumefaciens* in order to obtain a normal phenotype in regenerated plantlets, and transgenic plantlets were successfully obtained. Two selective systems, hygromycin (Hatanaka *et al.*, 1999) and the herbicide chlorsulfuron (Spiral *et al.*, 1999) proved efficient for selecting transformed embryogenic calli.

Transgenic coffee plants resistant to insects were obtained using this methodology (Leroy *et al.*, 1999 ; Spiral *et al.*, 1999). The leaf miner *Perileucoptera coffeella* is responsible of leaf degradation and, subsequently, yield decrease. Using *A. tumefaciens*, Spiral *et al.* (1999) successfully transferred the *Bacillus thuringiensis* *cry1Ac* gene conferring insect resistance into *C. canephora* and *C. arabica* genotypes. Efficiency of transformation varied depending on the genotype tested, the *arabica* genotypes being less amenable to embryo regeneration. Molecular characterization of transformed plants showed that 69% of them carried a unique copy of T-DNA, and Cry1Ac protein expression in leaves was obtained for 18 of 23 plantlets tested (Leroy *et al.*, 1999). Bioassays conducted using two leaf miner species showed the Cry1Ac protein conferred resistance to transgenic plants (Leroy *et al.*, 1999). Three different levels of resistance could be measured, with some highly resistant plants, slightly susceptible and fully susceptible plants. Agronomic evaluation and the insect resistance of the regenerated plants have still to be assessed in field trials.

7. Conclusions

The development of coffee varieties resistant to the major pests and diseases is a very challenging task that would considerably increase the sustainability of coffee production. The actual chemical control approach is very expensive, accounting for a large part of the production cost. Additionally, the long-term use of pesticides has negative effects on both human and environmental health.

Considerable knowledge on the genetic organisation of the *Coffea* genus and species has been acquired during this last decade owing to the use of molecular markers. Coffee research is now entering into the genomic and proteomic age and it is expected that numerous data will be soon generated. Furthermore, ongoing technological developments, including automation, allele-specific diagnostics and DNA chips, will make molecular approaches based on large-scale screening much more powerful and effective. All together, these advances have opened new perspectives in coffee breeding not only for resistance to parasites but also for improving other traits of importance such as resistance to abiotic stresses (frost, long-dry season..), biochemical and physical bean characteristics, and plant features adapted to the mechanical harvesting.

References

- Agwanda, C., Lashermes, P., Trouslot, P., Combes, M.C., and Charrier, A. (1997) Identification of RAPD markers for resistance to Coffee Berry Disease, *Colletotrichum kahawae*, in Arabica coffee, *Euphytica* **97**: 241-248.
- Anthony, F., Bertrand, B., Quiros, O., Lashermes, P., Berthaud, J. and Charrier, A. (2001) Genetic diversity of wild coffee (*Coffea arabica* L.) using molecular markers, *Euphytica* **118**: 53-65.
- Anzueto, F., Bertrand, B., Sarah, J.L., Eskes, A.B., and Decazy, B. (2001) Resistance to *Meloidogyne incognita* in Ethiopian *Coffea arabica* accessions, *Euphytica* **118**: 1-8.
- Barre, P., Akaffou, S., Louarn, J., Charrier, A., Hamon, S. and Noirot, M. (1998) Inheritance of caffeine and heteroside contents in an interspecific cross between a cultivated coffee species *Coffea liberica* and a wild species caffeine free *C. pseudozangebariae*, *Theor. Appl. Genet.* **96**: 306-311.
- Bennett, M.D. and Leitch, I.J. (1995) Nuclear DNA amounts in angiosperms, *Annals of Botany* **76**: 113-176.
- Bertrand, B., Anzueto, F., Pena, M., Anthony, F., and Eskes, A.B. (1995) Genetic improvement of coffee resistance to root-knot nematodes (*Meloidogyne* sp.) in Central America, *Proceedings of the XVI Scientific Colloquium on Coffee*, ASIC, Kyoto, Japan, ASIC, Paris, pp. 630-636.
- Bertrand, B., Aguilar, G., Santacero, R. and Anzueto, F. (1999) El mejoramiento genético en América Central, in B. Bertrand and B. Rapidel (eds.), *Desafíos de la caficultura centroamericana*, IICA, San José, pp. 407-456.
- Bertrand B., Anthony F., Lashermes P. (2001) Breeding for resistance to *Meloidogyne exigua* of *Coffea arabica* by introgression of resistance genes of *Coffea canephora*. *Plant Pathology* **50** : 637-644.
- Bettencourt, A.J. and Rodrigues, C.J. Jr. (1988) Principles and practice of coffee breeding for resistance to rust and other disease, in R.J. Clarke and R. Macrae (eds.), *Coffee vol.4: Agronomy*, pp 199-234 Elsevier Applied Science, London
- Beynon, S.M., Coddington, A., Lewis, B.G., Varzea, V. (1995) Genetic variation in the coffee berry disease pathogen, *Colletotrichum kahawae*, *Physiological and Molecular Plant Pathology* **46**:457-470
- Bolvenkel, E.T., Buckley and Eijgendaal, C. (1993) Report on a mission for the International Trade Centre. Dissemination Activities for Coffee – an exporters guide (Project Nos INT/50/33 – INT/24/64 INT/84/01)
- Bridson, D.M., and Verdcourt, B. (1988) Flora of tropical East Africa - *Rubiaceae* (part 2), 227 p, Polhill R.M. (ed), Balkema, Brookfield, Rotterdam.
- Brown, J.K.M. (1995) Recombination and selection in populations of plant pathogens, *Plant Pathology* **44**:279-293
- Brown, J.K.M. (1996) The choice of molecular markers methods for population genetic studies of plant pathogens, *New Phytol.* **133**:183-195
- Brygoo, Y., Caffier, V., Carlier, J., Fabre, J-V., Fernandez, D., Giraud, T., Mourichon, X., Neema, C., Notteghem, J-L., Pope, C., Tharreau, D. and Lebrun, M-H. (1998) Reproduction and population structure in phytopathogenic fungi, in P. Bridge, Y. Couteaudier and J. Clarkson (eds.) *Molecular variability of fungal pathogens*, CAB International, Oxon, pp. 133-148.
- Campos, V.P., Sivapalan, P. and Gnanapragasam, N.C. (1990) Nematode parasites of coffee, coca, and tea. In M. Luc, R.A. Sikora and J. Bridge, eds. *Plant-parasitic nematodes in subtropical and tropical agriculture*, CAB International, London, pp 113-126.
- Carneiro, R.M.D.G., Carneiro, R.G., Abrantes, I.M.O., Santos, M.S.N. and Almeida, M.R.A. (1996a) *Meloidogyne paranaensis* n. sp. (Nemata: Meloidogynidae) a root-knot nematode parasitizing coffee from Brazil, *Journal of Nematology*, **28**, 177-189.
- Carneiro, R.M.D.G., Almeida, M.R.A. and Carneiro, R.G. (1996b) Enzyme phenotypes of Brazilian isolates of *Meloidogyne* spp, *Fundamental and Applied Nematology*, **19**, 555-560.
- Carvalho, A. (1952) Taxonomia de *Coffea arabica* L. Caracteres morfológicos dos haploides, *Bragantia* **12**: 201-212.
- Castagnone-Sereno, P., Vanlerberghe-Masutti, F. and Leroy, F. (1994) Genetic polymorphism between and within *Meloidogyne* species detected with RAPD markers, *Genome*, **37**, 904-909.
- Cenis, J.L. (1993) Identification of four major *Meloidogyne* spp. by random amplified polymorphic DNA (RAPD-PCR), *Phytopathology* **83**:76-78
- Charrier, A. and J. Berthaud (1985) Botanical classification of coffee, in: Clifford MN and Wilson KC (eds.), *Coffee: Botany, biochemistry and production of beans and beverage*. Croom Helm, London, Sydney, pp 13-47.

- Choi, D., Kim, H.M., Yun, H.K., Park, J.A., Kim, W.T. and Bok, S.H. (1996) Molecular cloning of a metallothioneine-like gene from *Nicotiana glutinosa* L. and its induction by wounding and tobacco mosaic virus infection, *Plant Physiol.* **112**:353-359.
- Clough, S.J, Fengler, K.A., Yu, I.C., Lippok, B., Smith, R.K., Bent, A.F. (2000) The Arabidopsis dnd1 "defense, no death" gene encodes a mutated cyclic nucleotide-gated ion channel, *Proc. Natl. Acad. Sci. USA* **16**:9323-8.
- Combes, M.C., Andrzejewski, S., Anthony, F., Bertrand, B., Rovelli, P., Graziosi, G. and Lashermes, P. (2000) Characterisation of microsatellite loci in *Coffea arabica* and related coffee species, *Molecular Ecology* **9**: 1178-1180.
- Cros, J., Combes, M.C., Chabrilange, N., Duperray, C., Monnot des Angles, A. and Hamon, S. (1995) Nuclear DNA content in the subgenus *Coffea* (Rubiaceae): inter- and intra-specific variation in African species, *Can. J. Bot.* **73**: 14-20.
- Cros, J., Combes, M.C., Trouslot, P., Anthony, F., Hamon, S., Charrier, A., and Lashermes, P. (1998) Phylogenetic relationships of *Coffea* species: new evidence based on the chloroplast DNA variation analysis, *Molecular Phylogenetics and Evolution* **9**: 109-117.
- Dussert, S., Lashermes, P., Anthony, F., Montagnon, C., Trouslot, P., Combes, M.C., Berthaud, J., Noirot, M. and Hamon, S. (1999) Le caféier *Coffea canephora*. in *Diversité génétique des plantes tropicales cultivées*, Hamon P, Seguin M, Perrier X, Glaszmann JC (eds), CIRAD, Paris, pp. 175-194.
- Fernandez, D., Noir, S., Bon, M.-C., Combes, M.-C., Silva, M.C., Guerra-Guimaraes, L., Anthony, F., Bertrand, B. and Lashermes, P. (2001) Molecular physiology and genetics of coffee resistance to parasites. *Proceedings of the XIX Scientific Colloquium on Coffee*, ASIC, Trieste, Italy, Asic (eds.), Paris, *in press*
- Gonçalves, W., and Pereira, A.A. (1998) Resistência do cafeeiro a nematóides. IV. Reação de cafeeiros derivados do Híbrido de Timor a *Meloidogyne exigua*, *Nematologia Brasileira* **22**: 39-50.
- Guerrero Filho, O., Silvarolla, M.B. and Eskes, A.B. (1999) Expression and mode of inheritance of resistance in coffee to leaf miner *Perileucoptera coffeella*, *Euphytica* **105**: 7-15.
- Guy, C.L. and Li, Q.B. (1998) The organization and evolution of the spinach stress 70 molecular chaperone gene family, *Plant Cell* **10**:539-556.
- Hatanaka, T., Choi, Y.E., Kusano, T., Sano, H. (1999) Transgenic Plants of Coffee *Coffea Canephora* From Embryogenic Callus Via Agrobacterium Tumefaciens-Mediated Transformation, *Plant Cell Reports* **19**:106-110
- Hillel, J., Schaap, T., Haberfeld, A., Jeffreys, A.J., Plotzky, Y., Cahaner, A. and Lavi, U. (1990) DNA fingerprints applied to gene introgression in breeding programs, *Genetics* **124**: 783-789.
- Holguin Melendez, F. (1993) Contribution à la recherche d'une résistance durable du caféier (*Coffea* spp.) à la rouille orangée (*Hemileia vastatrix* Berk. et Br.). Etude de la variabilité du pathogène, PhD thesis, Univ. Montpellier II, 172 pp
- Hospital, F., Chevalet, C., Mulsant, P. (1992) Using markers in gene introgression breeding programs, *Genetics* **132**: 1199-1210.
- Kushalappa, A.C. and Eskes, A.B. (1989) Advances in coffee rust research, *Annu. Rev. Phytopathol.* **27**:503-531
- Ky, C.L., Barre, P., Lorieux, M., Trouslot, P., Akaffou, S., Louarn, J., Charrier, A., Hamon, S. and Noirot, M. (2000) Interspecific genetic linkage map, segregation distortion and genetic conversion in coffee (*Coffea* sp.), *Theor. Appl. Genet.* **101**: 669-676.
- Lashermes, P., Couturon, E., Moreau, N., Paillard, M. and Louarn, J. (1996c) Inheritance and genetic mapping of self-incompatibility in *Coffea canephora* Pierre, *Theor. Appl. Genet.* **93**: 458-462.
- Lashermes, P., Cros, J., Combes, M.C., Trouslot, P., Anthony, F., Hamon, S. and Charrier, A. (1996b) Inheritance and restriction fragment length polymorphism of chloroplast DNA in the genus *Coffea* L., *Theor. Appl. Genet.* **93**:626-632.
- Lashermes, P., Combes, M.C., Robert, J., Trouslot, P., D'Hont, A., Anthony, F. and Charrier, A. (1999) Molecular characterisation and origin of the *Coffea arabica* L. genome, *Mol. Gen. Genet.* **261**:259-266.
- Lashermes, P., Combes, M.C., Trouslot, P. and Charrier, A. (1997) Phylogenetic relationships of coffee-tree species (*Coffea* L.) as inferred from ITS sequences of nuclear ribosomal DNA, *Theor. Appl. Genet.* **94**: 947-955.
- Lashermes, P., Trouslot, P., Anthony, F., Combes, M.C. and Charrier, A. (1996a) Genetic diversity for RAPD markers between cultivated and wild accessions of *Coffea arabica*, *Euphytica* **87**: 59-64.
- Lashermes, P., Andrzejewski, S., Bertrand, B., Combes, M.C., Dussert, S., Graziosi, G., Trouslot, P. and Anthony, F. (2000b) Molecular analysis of introgressive breeding in coffee (*Coffea arabica* L.), *Theor. Appl. Genet.*, **100**: 139-146.

- Lashermes, P., Paczek, V., Trouslot, P., Combes, M.C., Couturon, E. and Charrier, A. (2000a) Single-locus inheritance in the allotetraploid *Coffea arabica* L. and interspecific hybrid *C. arabica* x *C. canephora*, *J. of Heredity* **91**: 81-85.
- Lashermes, P., Combes, M.C., Topart, P., Graziosi, G., Bertrand, B., and Anthony, F. (2000c) Molecular breeding in coffee (*Coffea arabica* L.), in Sera, Soccol, Pandey and Roussos, (eds.), *Coffee Biotechnology and Quality*, Kluwer Academic Publisher, Dordrecht, pp. 134-146.
- Lashermes, P., Combes, M.C., Prakash, N.S., Trouslot, P., Lorieux, M. and Charrier, A. (2001) Genetic linkage map of *Coffea canephora*: effect of segregation distortion and analysis of recombination rate in male and female meioses, *Genome*, **44** : 589-596.
- Leroy, T., Philippe, R., Royer, M., Frutos, R., Duris, D., Dufour, M., Jourdan, I., Lacombe, C. and Fenouillet, C. (1999) Genetically modified coffee plants expressing the *Bacillus thuringiensis cryIAc* gene for resistance to leaf miner, *Proceedings of the XVIII Scientific Colloquium on Coffee*, ASIC, Helsinki, Finland, Asic (eds.), Paris, 332-338
- Leung, H., Nelson, R.J., Leach, J.E. (1993) Population structure of plant pathogenic fungi and bacteria, *Advances in Plant Pathology* **10**:157-205
- Lopez, R. and Salazar, L. (1989) *Meloidogyne arabicida* sp. (Nematoda: Heteroderidae) nativo de Costa Rica: un nuevo y severo patógeno del café, *Turrialba*, **39**, 313-323
- Manga B. (1999) Etude de l'antracnose des baies du caféier causée par *Colletotrichum kahawae*, PhD thesis, Univ. Montpellier II, 165 pp.
- Marraccini, P., Deshayes, A., Petiard, V., Rogers, W.J. (1999) Molecular Cloning of the Complete 11s Seed Storage Protein Gene of *Coffea Arabica* and Promoter Analysis in Transgenic Tobacco Plants, *Plant Physiology and Biochemistry* **37**:273-282
- McDonald, B.A. (1997) The population genetics of fungi: Tools and techniques, *Phytopathology* **87**:448-453
- Mettulio, R., Rovelli, P., Anthony, F., Anzueto, F., Lashermes, P. and Graziosi, G. (1999) Polymorphic microsatellites in *Coffea arabica*. , *Proceedings of the XVIII Scientific Colloquium on Coffee*, Helsinki (Finlande), pp. 344-347.
- Melchinger, A.E. (1990) Use of molecular markers in breeding for oligogenic disease resistance, *Plant Breed.*, **104**: 1-19.
- Milgroom, M.G., Fry, W.E. (1997) Contributions of population genetics to plant disease epidemiology and management, *Advances in Botanical Research* **24**:1-30
- Moreno, G., Cortina, H., Alvarado, G. and Gaitan, A. (2000) Utilización de los recursos genéticos de café en el programa de mejoramiento genético de *C. arabica*, en Colombia. Atelier sur l'amélioration durable du caféier *Coffea arabica*, 29-30 août 2000, CATIE, Turrialba (Costa Rica), pp 33-38.
- Nandris, D., Kohler, F., Fernandez, D., Lashermes, P., Rodrigues, J. and Pellegrin, F. (1998) Coffee pathosysteme modelling : 2- assessment of host and pathogen biodiversities. 7th International Congress of Plant Pathology, Edinburgh, 9-16 August, Scotland.
- Noir, S., Combes, M.C., Anthony, F. and Lashermes, P. (2001) Origin, diversity and evolution of NBS disease-resistance gene homologues in coffee trees (*Coffea* L.), *Molecular Genetics and Genomics* **265** : 654-662.
- Orozco-Castillo, C., Chalmers, K.J., Waugh, R. and Powell, W (1994) Detection of genetic diversity and selective gene introgression in coffee using RAPD markers, *Theor. Appl. Genet.* **87**: 934-940.
- Orozco-Castillo, C., Chalmers, K.J., Powell, W. and Waugh, R. (1996) RAPD and organelle specific PCR re-affirms taxonomic relationships within the genus *Coffea*. *Plant Cell Reports* **15**: 337-341
- Paillard, M., Lashermes, P. and Pétiard, V. (1996) Construction of a molecular linkage map in coffee, *Theor. Appl. Genet.* **93**: 41-47.
- Paterson, A.H., Tanksley, S.D. and Sorrells, M.E. (1991) DNA markers in plant improvement, *Advances in Agronomy* **46**: 39-90.
- Raina, S.N., Mukai, Y. and Yamamoto, M. (1998) *In situ* hybridisation identifies the diploid progenitor of *Coffea arabica* (Rubiaceae), *Theor. Appl. Genet.* **97**: 1204-1209.
- Rani, V., Singh, K.P., Shiran, B., Nandy, S., Goel, S., Devarumath, R.M., Sreenath, H.L., and Raina, S.N. (2000) Evidence for new nuclear and mitochondrial genome organizations among high-frequency somatic embryogenesis-derived plants of allotetraploid *Coffea arabica* L. (Rubiaceae), *Plant Cell Rep.* **19**: 1013-1020.
- Richter, T.E. and Ronald, P.C. (2000) The evolution of disease resistance genes, *Plant Molecular Biology* **42**: 195-204.
- Rodrigues, C.J.Jr., Bettencourt, A.J. and Rijo, L. (1975) Races of the pathogen and resistance to coffee rust, *Annu. Rev. Phytopathol.* **13**: 49

- Rovelli, P., Mettullio, R., Anthony, F., Anzueto, F., Lashermes, P. and Graziosi, G. (2000) Microsatellites in *Coffea arabica* L., in Sera, Soccol, Pandey and Roussos (eds.), *Coffee Biotechnology and Quality*, Kluwer Academic Publisher, Dordrecht, pp. 123-133
- Sessa, G. and Martin, G.B. (2000) Signal recognition and transduction mediated by the tomato Pto kinase: a paradigm of innate immunity in plants, *Microbes and infection* **2**:1591-1597
- Semblat, J.P., Wajnberg, E., Dalmasso, A., Abad, P. and Castagnone-Sereno, P. (1998) High-resolution DNA fingerprinting of parthenogenetic root-knot nematodes using AFLP analysis, *Mol. Ecol.* **7**:119-125.
- Sera T. (2000) Genetics of coffee resistance to nematode *Meloidogyne paranaensis*. Seminar on "Mejoramiento sostenible del café Arabica por los recursos genéticos, asistido por los marcadores moleculares, con énfasis en la resistencia a los nemátodos". CATIE, Costa Rica, pp. 61-65.
- Spiral, J. and Pétiard, V. (1993) Développement d'une méthode de transformation appliquée à différentes espèces de caféier et régénération de plantules transgéniques. *Proceedings of the XV Scientific Colloquium on Coffee*, ASIC, Montpellier, France, Asic (eds.), Paris, pp.115-122
- Spiral, J., Leroy, T., Paillard, M. and Pétiard, V. (1999) Transgenic coffee (*Coffea* species), in Y.P.S. Bajaj (ed.), *Biotechnology in Agriculture and Forestry, Vol. 44, transgenic Trees*, Springer-Verlag, Berlin, Heidelberg, pp. 55-76
- Sreenivasaprasad, S., Brown, A.E., Mills, P.R. (1993) Coffee berry disease pathogen in Africa : genetic structure and relationship to the group species *Colletotrichum gloeosporioides*, *Mycol. Res.* **97**:995-1000
- Stam, P., and Zeven, A.C. (1981) The theoretical proportion of the donor genome in near-isogenic lines of self-fertilizers bred by backcrossing., *Euphytica* **30**: 227-238
- Sugiyama, M., Matsuoka, C. and Takagi, T. (1995) Transformation of coffee with *Agrobacterium rhizogenes*, *Proceedings of the XVI Scientific Colloquium on Coffee*, ASIC, Kyoto, Japan, Asic (eds.), Paris, pp.853-859
- Tedesco, N.S., Sakiyama, N.S., Zambolim, L., Teixeira, T.A., Periera, A. and Sakiyama, C.C.H. (1999) Mapeamento de gene de resistencia a ferrugem-do-café com marcador RAPD, III International Seminar on Biotechnology in the coffee agroindustry (SIBAC), Londrina (Brasil), p. 185.
- Van Boxtel, J., Berthouly, M., Carasco, C., Dufour, M. and Eskes, A. (1995) Transient expression of beta-glucuronidase following biolistic delivery of foreign DNA into coffee tissues, *Plant Cell Reports* **14**:748-752
- Van der Vossen, H.A.M. and Walyaro, D.J. (1980) Breeding for resistance to Coffee Berry Disease in *Coffea arabica* L. II. Inheritance of resistance., *Euphytica* **29**: 777-791
- Van Loon, L.C. and Van Strien, E.A. (1999) The families of pathogenesis-related proteins, their activities, and comparative analysis of PR-1 type proteins, *Physiol. Mol Plant Pathol.* **55**:85-97.
- Waller, J.M., Bridge, P.D., Black, R. and Hazika, G. (1993) Characterization of the coffee berry disease pathogen, *Colletotrichum kahawae* sp. nov., *Mycol. Res.* **97**:989-994
- Whitbred, J.M., Schuler, M.A. (2000) Molecular Characterization of Cyp73a9 and Cyp82a1 P450 Genes Involved in Plant Defense in Pea, *Plant Physiology* **124**:47-58

CONSTRUCTION AND USE OF GENETIC MAPS IN CEREALS

M. MOTTO, P. AJMONE MARSAN

*Istituto Sperimentale per la Cerealicoltura, Sezione di Bergamo
Via Stezzano 24, 24126 Bergamo, Italy*

1. INTRODUCTION

Breeding cereal crops has been remarkably successful in the improvement of both qualitative and quantitative traits that affect agronomic performance and consumer preferences. In wheat, rice, and maize for example about 50% of the increase in the yield over the past decades has been attributed to improved varieties; the remainder derives from greater and more efficient use of fertilizers and better crop management.

Recent advances in genetics, cell and molecular biology, particularly recombinant-DNA technology provide new opportunities for manipulation of cereal genome to meet the demands of breeders. In this context, the development of molecular marker technology and consequent identification of marker loci linked to important agronomic traits have also created exciting new opportunities for plant breeders. One of the priorities of plant genome mapping is the identification of genes associated with economically important traits and the use of this information to further improve crops. The value of molecular markers for the development of linkage maps and their use in the analysis of economically important traits has been demonstrated in several crops including cereals (reviewed by Mohan *et al.*, 1997).

In the past decade the creation of genetic maps has been the foundation for this new plant breeding approach. These maps provide expectation for gene tagging (i.e. identification of markers tightly linked to major genes), analysis of quantitative trait loci (QTLs), map-based gene cloning, and characterisation of genetic variation in germplasm collections. In addition, marker-assisted selection (MAS) provides the potential for improving selection efficiency for agronomic traits. In this chapter we describe how genetic markers are used in: i) map construction; ii) identification of QTLs, and iii) defining the potential use of genetic maps for MAS in cereals.

2. GENETIC MARKERS

Successful application of genetic markers in crop improvement may depend on the availability of highly saturated linkage maps and on the ability to analyse large populations efficiently and inexpensively. Morphological and isozyme markers were the first used to construct linkage maps in several cereals (O'Brien, 1993) and in marker assisted selection programmes. For instance, in maize the allelic frequencies at a number of enzyme loci were modified by recurrent selection based on grain yield (Stuber *et al.*, 1980). Accordingly, the selection based only on

manipulation of allelic frequencies at enzyme loci induced yield increase in maize (Frei *et al.*, 1986). Using isozymes, genes or gene clusters controlling a large number of quantitative traits were selected and mapped in maize. Genes affecting QTLs appeared to be distributed throughout the maize genome with some regions affecting a greater number of traits than others (Edwards *et al.*, 1987; Stuber *et al.*, 1987). Using 17-20 marker loci, these authors found that the cumulative effects of marker-linked genome regions explained from 8 to 40% of the phenotypic variation for 25 traits. Moreover, a single trait was associated to a minimum of 7 and to a maximum of 19 isozyme markers.

Although, the information on these conventional maps is important to know the location of genes corresponding to phenotypic traits, their usefulness is limited by the low number of morphological markers available for crop improvement programmes. The advent of molecular (DNA) marker techniques has provided virtually unlimited supply of phenotypically neutral markers, opening new possibilities and horizons for genetic mapping (Stuber, 1995). Technological developments during the last 20 years have expanded the range of DNA polymorphism assays employed in genome mapping of cereals, the first of which was restriction fragment length polymorphisms (RFLPs; Botstein *et al.*, 1980). However, RFLPs have some drawbacks, like the use of radioactive labelling, the laboriousness of the technique, the large amount of DNA required, and the unamenability to automation, which have prompted the search for more efficient marker systems.

In the past few years, a generation of DNA marker systems based on the polymerase chain reactions (PCR) has emerged. These marker systems, such as simple sequence repeats (SSRs; Tautz, 1989), random amplified polymorphic DNAs (RAPDs; Williams *et al.*, 1990), and the AFLP[®] assay (Vos *et al.*, 1995), rely on the exponential amplification of subsets of the total amount of DNA sequence variation in a genome using either random or specific oligonucleotide priming sets.

In barley and maize comparative analyses of marker efficiency for RFLPs, RAPDs, SSRs, and AFLPs indicate that these assays differ in the amount of polymorphism detected (Russell *et al.*, 1997; Pejic *et al.*, 1998). The information content, measured by the expected heterozygosity and the average number of alleles per locus, was higher for SSRs, while the lowest level of polymorphism was obtained with AFLPs. However, AFLPs were the most efficient marker system because of their ability to reveal several bands in a single amplification. In fact, the assay efficiency index was more than ten-fold higher for AFLPs compared with other methods. An example of a comparison of AFLP information content with other molecular marker techniques among a set of maize inbred lines is shown in Table 1 (Pejic *et al.*, 1998). Other information indicates that AFLP markers are, generally, distributed throughout the genome, although clustering of markers in centromeric regions has also been reported (Qi *et al.*, 1998; Castiglioni *et al.*, 1999; Vuylsteke *et al.*, 1999). There is also evidence that AFLP markers lie outside regions that are heavily populated with RFLPs (Waugh *et al.*, 1997; Castiglioni *et al.*, 1999).

The necessity of a high-throughput low-cost genotyping technology in genome mapping is also related to automation. DNA polymorphisms that lend themselves very well to automated analysis on a very large scale, e.g. with DNA array technology, are single nucleotide polymorphisms (SNPs). Maize SNP project at DuPont and Pioneer aimed at evaluating the potential of SNPs as the next generation of genetic markers suggests that this approach is capable of very high throughput genotyping for genetic mapping (Bhatramakki *et al.*, 2000a). These authors, by aligning the sequences of amplification products from 20 loci randomly distributed in the maize genome, found a level of polymorphisms as high as 1 single nucleotide change per 70 bp. Therefore, biochips containing plant SNPs are expected to enter the field of molecular plant breeding soon.

TABLE 1. Level of polymorphism and comparison of informativeness obtained with RFLP, RAPD, SSR and AFLP markers in 33 maize inbred lines (adapted from Pejic *et al.*, 1998)

Parameters	Marker system			
	RFLP	RAPD	SSR	AFLP
Number of assay units	53 (p/e) ^a	25 (p) ^b	27 (p/p)	6 (p/c) ^d
Number of polymorphic bands	253	90	183	232
Number of loc	53	90 ^c	27	232 ^c
Average number of alleles per locus	4.8	2.0	6.8	2.0
Expected heterozygosity	0.63	0.36	0.72	0.34
Effective number of allele per locus	3.2	1.6	4.4	1.6
Assay efficiency index	3.2	5.8	4.4	61.9

^a (p/e) = (probe/enzymes); ^b (p) = (primers); ^c (p/p) = (primer pairs);

^d (p/c) = (primer combinations); ^e theoretical maximum number of loci

3. MOLECULAR MAPS AND THEIR USE FOR LOCATING GENES

In the past few years, the density of maps of all major cereals have increased up to 50% and mapping is progressing in species as rye, oat, sorghum, and millet. Most maps were mainly constructed on the basis of RFLP markers. In recent years, SSR, AFLP, and RAPD markers were also used in genome mapping of cereals. However, current data are available through conferences and proceedings, in species-specific newsletters, such as the Maize Genetics Cooperative Newsletters and the Rice Genetics Newsletters, through coordinated groups or via computerised databases (e.g. Probe-USDA Plant Genome Research Programme Newsletter, for information see the World Wide Web site: <http://ars-genome.cornell.edu/grasses.html>).

3.1. Maize

Maize research has a long tradition in the area of gene mapping. This plant has a genetic map encompassing over 600 gene loci, many of which are expressed as seedling or kernel factors (Coe *et al.*, 1995). The cytogenetics of maize is also well developed and has played a pivotal role in defining many aspects of chromosome behaviour, in addition to equipping the maize geneticists with a rich resource of tools for addressing genetic problems (Carlson, 1988). Recessive genes can be mapped to 19 of the 20 chromosome arms using B-A translocations and dominant genes can be localised using waxy-marked translocation stocks.

The application of RFLPs to genetic mapping in maize has helped to further characterise the maize genome. This is due, at least in part, to the fact that maize is highly polymorphic (Evola *et al.*, 1986). Using pooled data from several F₂ populations, Helentjaris *et al.* (1986) constructed the first published RFLP map of maize. Maps based on publicly available maize RFLP probes were further presented by Coe *et al.* (1987) and Burr *et al.* (1988) that probed random genomic clones on an F₂ population and on a set of recombinant inbred lines (RILs), respectively. Gardiner *et al.* (1993) published an updated version of the map published by Coe *et al.* (1987) on the basis of an immortalized version of the same F₂ (IF₂) population. Weber and Helentjaris (1989) used RFLPs to analyse the progeny of crosses between B-A translocation lines and other inbreds, to link cytological information to the

molecular map. Beavis *et al.* (1992) developed a mapping population based on random mating that provided improved resolution compared to F_2 or RIL populations of similar size. A composite map based on four mapping populations, containing 275 loci representing both expressed sequence tagged sites (ESTs) and anonymous sequences was published by Causse *et al.* (1996). More recently, Davis *et al.* (1999) have constructed a 1,736-locus maize genome map containing 1,156 loci probed by cDNA, 545 probed by random genomic clones, 16 by SSRs, 14 by isozymes, and 5 by anonymous clones. Sequence information is available for 56% of the loci, with 66% of the loci sequenced having a function assigned. A total of 596 new ESTs from a B73 library of 5-wk-old shoots were located on this map. In addition the map contains 237 loci probed by barley, oat, wheat, rice, and *Tripsacum* clones, which serve as grass genome reference points in comparisons between maize and other grass maps. Ninety core markers selected for low copy number, high polymorphism, and even spacing along the chromosomes delineate 100 bins¹ on the map. The average bin size is 17 cM. Use of bin assignments enables comparison among different maize mapping populations and experiments, including those involving cytogenetic stocks, mutants, and QTLs. Integration of non-maize markers in the map extends the resources available for gene discovery beyond the boundaries of maize mapping information into map, sequence, and phenotype information from other grass species.

SSR markers were also utilised to map genes in maize (Taramino and Tingey, 1996). Senior *et al.* (1997) published the first maize linkage map based upon SSR markers. Vuylsteke *et al.* (1999) have developed two high-density AFLP linkage maps of maize based on a B73/Mo17 RIL population and a D32/D145 IF_2 population. A total of 1,539 and 1,355 AFLP markers have been mapped in the two populations, respectively. Although the number of AFLPs mapped was lower in the IF_2 map, most IF_2 linkage groups were longer than the homologous RIL groups. In addition, the total IF_2 map length (1,376 cM) was more comparable than RIL (1,178 cM) to the lengths reported in other maps. In both maps, the position of RFLP, SSR, and isozyme markers were consistent with those previously published maize maps. These authors have also shown, by map comparison that, despite some rearrangements, most of the AFLP markers in common between populations, map at similar positions. Changes in map order occur mainly in marker-dense regions. These marker-dense regions co-localize well with the putative centromeric regions of the maize chromosomes. Similarly, Castiglioni *et al.* (1999) exploited the AFLP technique to saturate an existing RFLP linkage map derived from a 229- F_2 maize mapping population from the cross B73/A7. These workers, by using restriction enzymes differing in methylation sensitivity, were able to detect 1,568 visible bands and map 246 AFLP markers covering 2,057 cM.

A number of other research groups have produced maize DNA maps with the intent of mapping quantitative traits relatively to the molecular markers. These maps provide a foundation for numerous basic and applied investigations, including studies of gene organisation, gene and genome evolution, targeted cloning, and dissection of complex traits. With the use of these DNA molecular maps, QTLs have been identified and characterised for several traits including grain yield and its components (Edwards *et al.*, 1992; Stuber *et al.*, 1992; Graham *et al.*, 1995; Austin and Lee, 1996a, b; Ajmone Marsan *et al.*, 1995, 2000), forage yield and quality traits (Lübberstedt *et al.*, 1997), morphological and physiological traits, such as plant height, tassel shape, flowering time (Koester *et al.*, 1993; Schön *et al.*, 1994; Veldboom *et al.*, 1994; Ribaut *et al.*, 1996; Berke and Rocheford, 1999), herbicide tolerance (Sari Gorla *et al.*, 1997), low phosphorus stress tolerance (Reiter *et al.*,

¹ (bin locations are designed by an X, Y code, where X is the linkage group containing the bin and Y is the location of the bin value within the linkage group)

1991), thermotolerance (Frova and Sari-Gorla, 1993), and pollen competitive ability (Ottaviano *et al.*, 1991). Moreover, QTLs were identified for kernel protein, starch, and oil concentrations (Goldman *et al.*, 1993, 1994), and for leaves abscisic acid concentration (Lebreton *et al.*, 1995; Tuberosa *et al.*, 1998). In addition, several investigations have reported QTLs conferring insect resistance (Bohn *et al.*, 1996, 1997), silk maysin concentration and antibiosis to corn earworm (Byrne *et al.*, 1996), and QTLs conferring resistance to pathogens, like northern corn leaf blight (Freyemark *et al.*, 1993), southwestern corn rust (Holland *et al.*, 1998), grey leaf spot (Saghai-Marooft *et al.*, 1996), head smut (Lübberstedt *et al.*, 1999), anthracnose stalk rot (Jung *et al.*, 1994), common smut (Lübberstedt *et al.*, 1998), and virus resistance (Melchinger *et al.*, 1998; Pernet *et al.*, 1999). This technology has also been used to explore QTLs important for an understanding of the evolution of maize from teosinte (Doebley and Stec, 1991), for controlling *in vitro* growth and differentiation of plant tissue (Armstrong *et al.*, 1992).

Beavis (1994) has summarised data reported on QTLs for plant height and grain yield. The number of QTLs detected ranged from 2 to 8 per experiment and the estimated contribution to trait variation of any single QTL ranged from 5 to 25% of the phenotypic variability. The number of progenies evaluated in these experiments ranged from 100 to 400 for plant height and between 100 and 250 for grain yield; the type of progenies included F_2 derived lines evaluated *per se*, backcrossed to parental inbred line(s) or testcrossed to unrelated tester line(s). In addition, results from three independent experiments repeated in the same genetic background (B73/Mo17) revealed that QTLs identified were not consistent (Stubber *et al.*, 1992; Beavis *et al.*, 1994). A number of confounding factors such as different sets of markers, sources of parental lines, type of progeny, different sets of environments, sampling of progenies, heritability of the traits under investigation, and biometrical approach employed for QTL mapping have been reported as possible causes of this discrepancy.

In this context Ajmone Marsan *et al.* (2001) have carried out a study for the identification of QTLs for grain yield and grain-related traits of testcross progenies using an AFLP map, different testers, and cofactor analysis. In this study, maize elite inbred lines were crossed to produce 229 F_2 individuals which were genotyped with 66 RFLP and 246 AFLP marker loci. By selfing the F_2 plants 229 F_3 lines were produced and subsequently crossed to two inbred testers (T1 and T2). Each series of testcrosses was evaluated in field trials for grain yield, dry matter concentration, and test weight.

The results of this study showed that the efficiency of generating AFLP markers was substantially higher relative to RFLP markers, and that the speed at which AFLPs were generated showed a great potential for application in MAS. In addition, it was found that AFLP markers covered genomic regions left uncovered by RFLPs, in particular at telomeric regions previously almost devoided of RFLP markers. This increase of genome coverage afforded by the inclusion of AFLPs revealed new QTL locations for all the traits investigated and allowed to map telomeric QTLs with higher precision. The study has also provided an opportunity to compare simple (SIM) and composite interval mapping (CIM) for QTL analysis. The results indicated that CIM has greater potential in the detection of QTLs, and provides more precise and accurate estimates of QTL positions and effects than SIM. For all traits and both testers the authors detected a total of 36 QTLs, of which only 2 were in common between testers (Figure 1). This suggested that the choice of a tester for identifying QTL alleles to be used in improving an inbred is critical and that the expression of QTL alleles may be tester-specific.

Khavkin and Coe (1997) have analysed conjointly the map locations reported for genes for growth, development, and stress response. They found that these genes associate into 10-30 cM long functional clusters and are non-randomly distributed along the chromosomes. The majority of over 800 QTLs for plant architecture, growth and development *in vivo* and *in vitro*, grain yield, as the integer of growth, and ABA accumulation and effects also map within these clusters. The authors concluded that these clusters are the functional units of genes expressed in

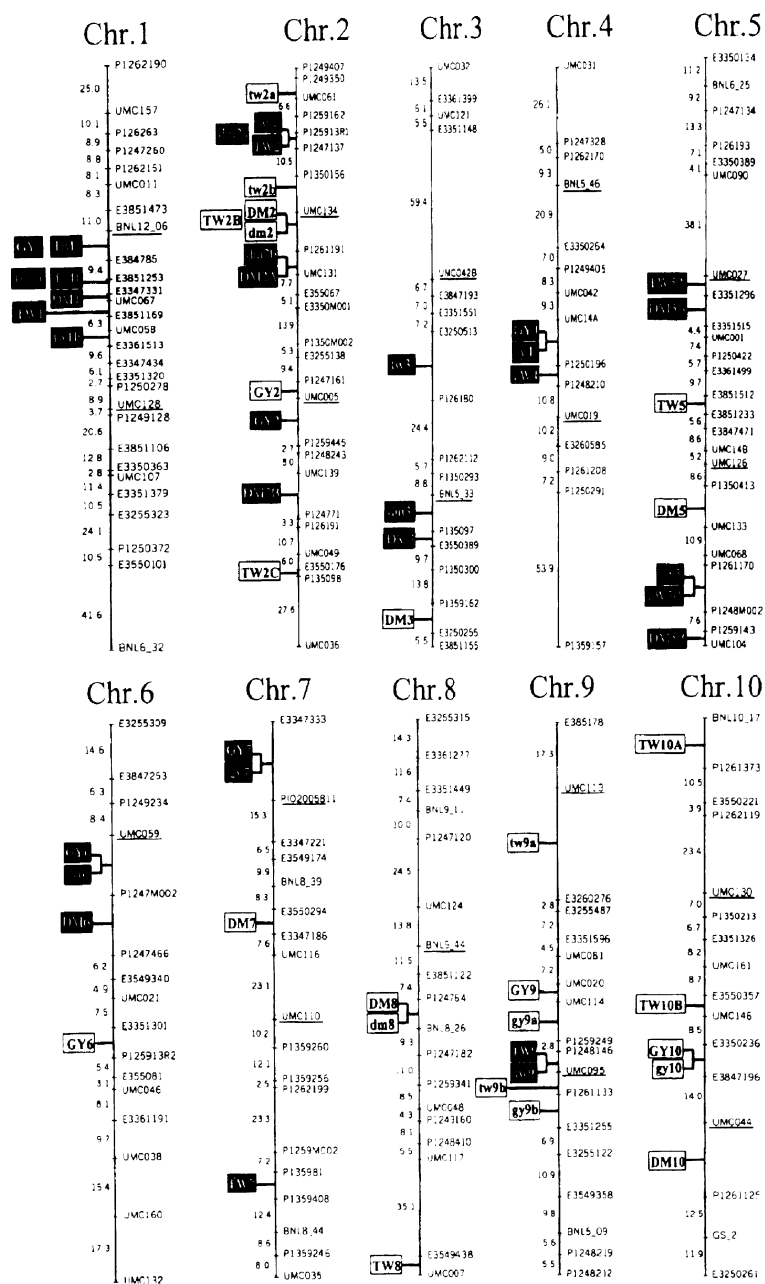


Figure 1

concert to contribute toward regulating plant development. Some of the plant responses to abiotic stress and the major QTLs for plant height, earliness and grain yield are suggested to be visible manifestation of developmental clusters.

3.2. Rice

Linkage maps were originally established with morphological mutants, but isozyme markers were subsequently located on the chromosomes (reviewed in Ishikawa *et al.*, 1991). McCouch *et al.* (1988) described the construction of the first RFLP map in rice. This map was constructed from an F_2 population derived from a cross between varieties representing the two major subspecies (*indica* and *japonica*) of cultivated rice. A second RFLP map of rice based on a different *indica/japonica* cross was reported by Saito *et al.* (1991). Moreover, Causse *et al.* (1994) described an RFLP map based on an interspecific backcross population involving *O. sativa* and *O. longistaminata*. The map contains 726 markers (mainly RFLPs) and spans 1,491 cM with an average interval size of 4.0 cM on the framework map, and 2.0 cM overall.

An additional map based on an F_2 population was developed at the Rice Genome Research Program in Japan by Harushima *et al.* (1998). This high resolution genetic map contained 2,275 DNA markers at an average interval of approximately 190 kb. The markers, distributed along 1,522 cM on 12 linkage groups, comprise 1,455 cDNAs, 265 genomic DNAs, 147 randomly amplified polymorphic DNAs (RAPD), and 88 other DNAs. cDNAs were derived from rice root and callus, analysed by single-run sequencing and searched for similarities with known proteins. Nearly 260 rice genes were newly identified and mapped, and genomic DNA and cloned RAPD fragments were also sequenced to generate SSRs.

Maheswaran *et al.* (1997) have exploited the AFLP technique to study the polymorphism, distribution, and inheritance of AFLP markers in a doubled haploid (DH) rice population derived from the cross IR64/Azucena. Using 20 primer combinations, they detected 945 AFLP bands, out of which 208 were polymorphic. All 208 AFLP markers were mapped and distributed over all 12 chromosomes already mapped in the same population with RFLP markers (Huang *et al.*, 1994). The map length of rice increased from 1,811 cM to 3,058 cM, and the average distance between markers was reduced from 13.4 cM to 8.9 cM. In addition, a molecular map of rice consisting of 231 AFLPs, 212 RFLPs, 86 SSRs, five isozyme loci, and two morphological mutant loci has been constructed using an F_{11} RIL population derived from two cultivated varieties (Cho *et al.*, 1998). All marker types were well distributed throughout the 12 chromosomes. The integrated map covered 1,814 cM, with an average interval size of 3.4 cM. This map is a cornerstone of the Korean Rice Genome Research Program and is being continuously refined through the addition of partially sequenced cDNA markers derived from an immature-seed cDNA library developed in Korea, and microsatellite markers developed at Cornell University. The population is also being used for QTL analysis and as the basis for marker-assisted variety development.

Figure 1. Linkage map of ten maize chromosomes based on the F_2 mapping population derived from the cross of inbred lines B73 and A7. To define chromosome regions, RFLP probes which localise centromeric regions on the reference map by Davis *et al.* (1998), and in common with our map, have been selected (underlined markers). Map distances, on the left side of the bars, are in centimorgans (cM) calculated using Haldane function. The position of QTLs identified with simple interval mapping (SIM) and composite interval mapping (CIM) are indicated in white and black boxes, respectively; small and capital letters reported within boxes refer to QTLs identified with tester T1 and T2, respectively. The presence of different QTLs on the same chromosome are indicated with the letters a,b,c, and A,B,C for SIM and CIM analysis, respectively.

GY: grain yield; DW: dry matter; TW: test weight (adapted from Ajmone Marsan *et al.*, 2001).

Altogether the rice maps provided the basis for locating a number of agronomically important genes *via* linkage to RFLP, AFLP and SSR markers, including both single genes and QTLs linked to blast resistance (Yu *et al.*, 1991; Wang *et al.*, 1994), insect resistance (McCouch and Tanksley 1991; Mohan *et al.*, 1993), bacterial blight resistance (Ronald *et al.*, 1992), photoperiod sensitivity (Mackill *et al.*, 1993), grain aroma (Ahn *et al.*, 1992), wide compatibility (Liu *et al.*, 1992; Zheng *et al.*, 1992), root traits (Ali *et al.*, 2000; Price *et al.*, 2000), thermosensitive genetic male sterility (Dong *et al.*, 2000), phosphorous deficiency tolerance (Ni *et al.*, 1998), bacterial leaf streak (Tang *et al.*, 2000), seed shattering and dormancy (Cai *et al.*, 2000), cell-membrane stability (Tripathy *et al.*, 2000), aluminium tolerance (Wu *et al.*, 2000), and the semi-dwarf character, *sd-1* (Cho *et al.*, 1994), among others.

3.3. Wheat

Constructing linkage maps of the wheat genomes have posed special problems. This is due to the fact that cultivated wheats have genomes containing a large number of repetitive DNAs and very limited level of polymorphism (Chao *et al.*, 1989; Cadalen *et al.*, 1997). Because of this reduced polymorphism, gene and genome mapping has required the use of populations derived from wide crosses. However, RFLPs have been assigned to wheat chromosome arms (Anderson *et al.*, 1992; Devey and Hart, 1993) and detailed RFLP linkage maps exist for hexaploid wheat (Chao *et al.*, 1989; Liu and Tsunewaki, 1991; Devos *et al.*, 1992; Xie *et al.*, 1993; Nelson *et al.*, 1995a,b,c; Van Deynze *et al.*, 1995a; Marino *et al.*, 1996) and to related species *Triticum tauschii* (Coss) Schmal. (Gill *et al.*, 1991; Lagudah *et al.*, 1991), and *T. monococcum* (Van Deynze *et al.*, 1995a).

T. monococcum was used to produce high-density RFLP maps that would complement the genetic maps of *T. aestivum*. In this respect, Dubcovsky *et al.* (1996) have developed a genetic map of the diploid wheat, *T. monococcum*, involving 335 markers, including RFLP markers, isozymes, seed storage proteins, rRNA, and morphological loci. It was also shown from this study that *T. monococcum* and barley linkage groups are remarkably conserved; they differ however by a reciprocal translocation involving the long arms of chromosomes 4 and 5, and by paracentric inversions in the long arm of chromosome 1 and 4; the latter is in a segment of chromosome arm 4L translocated to 5L in *T. monococcum*. The order of the markers in the inverted segments of the *T. monococcum* genome is the same as in the B and D genomes of *T. aestivum*. The *T. monococcum* map differs from the barley maps in the distribution of recombination within chromosomes. However, the total distance between the most distal common markers in all chromosomes was 714.2 cM in *T. monococcum* and 721.6 cM in *T. aestivum*.

Mapping many agronomically important genes or QTLs, a major goal in plant breeding requires informative markers in an intraspecific context. This is particularly true for MAS. RFLPs detected with single-copy genomic and cDNA clones are extremely powerful for comparative mapping approaches (Ahn *et al.*, 1993; Moore *et al.*, 1995; Sherman *et al.*, 1995; Yu *et al.*, 1996). However, they are only of limited use for intraspecific molecular analysis of agronomic traits because usually <10% of all RFLP loci are polymorphic in wheat. To overcome this drawback, Roder *et al.* (1998) have reported the development of highly polymorphic SSR markers using procedures optimised for the large wheat genome. Only 20% of these markers detected more than one locus. A total of 279 loci amplified by 230 primer sets were placed onto a genetic framework map composed of RFLPs previously mapped in the reference population, Opata 85/W7984, of the International Triticeae Mapping Initiative. By using this approach the previous authors mapped 65 microsatellites at a LOD > 2.5, and 214 microsatellites were assigned to the most likely intervals. In addition, 93 loci were mapped to

the A genome, 115 to the B genome, and 71 to the D genome. The markers were randomly distributed across linkage groups, with clustering in several centromeric regions.

The genetic linkage maps can be compared among crop plants using a common set of DNA probes. Comparison among the molecular maps of wheat, barley, *T. tauschii* and *T. monococcum* (Devos *et al.*, 1993; Nelson *et al.*, 1995a,b; Van Deynze *et al.*, 1995a) indicate that the orders of molecular markers in the linkage maps of these species detected with the same probes are essentially homosequential. Consequently, consensus maps were constructed to represent the linkage maps for the chromosomes of these species (Nelson *et al.*, 1995a,b; Van Deynze *et al.*, 1995a). The linkage or chromosome arm maps of more distantly related species such as wheat, rice, maize, and oat (Ahn and Tanksley, 1993; Ahn *et al.*, 1993; Devos *et al.*, 1994; Kurata *et al.*, 1994; Van Deynze *et al.*, 1995a, b, c) show significant conservation in the order of DNA markers detected by the same probes.

Introgression of resistance genes has been the method of choice for controlling leaf rust diseases in wheat (McIntosh *et al.*, 1995). Therefore, it is not surprising that rust resistance (e.g. *Lr29*, *Lr9*, *Lr47*), and powdery mildew resistant genes (e.g. *Pm1* and *Pm2*, *Mire*) were frequent targets in RFLP mapping studies (Hart *et al.*, 1993; Schachermayr *et al.*, 1994, 1995; Autrique *et al.*, 1995; Dedryver *et al.*, 1996; Nelson *et al.*, 1997; Sun *et al.*, 1997; Dubcovsky *et al.*, 1998; Chantret *et al.*, 2000; Helguera *et al.*, 2000). QTL analyses of agronomic traits (Anderson *et al.*, 1993), restoration of fertility (Ma and Sorrells, 1995), and end-use quality traits have been also reported from crosses between wheat and related species (Zanetti *et al.*, 1998) or in tetraploid wheats (Blanco *et al.*, 1998). Zanetti *et al.* (1998) described three major QTLs on chromosome 5A for protein content, Zeleny and falling number, and mentioned others on chromosome 1A [low-molecular-weight (LMW) glutenin subunit] and on the homeologous group 6 and 7 chromosomes. Blanco *et al.* (1998) found a major effect of the Gli-B1/Glu/B3 loci and significant effects of loci located on chromosomes 1AL, 3AS, 3BL, 5AL, 6AL, and 7BS on sedimentation volume. QTLs for soluble pentosans and associated components were mapped in two progenies of hexaploid wheats (Martinant *et al.*, 1998). In addition, Perretant *et al.* (2000) identified QTLs for three bread-making quality traits: hardness, protein content, and strength of the dough. For hardness, the cited authors confirmed a previously tagged major QTL on chromosome 5DS, and two additional minor QTLs were found on chromosome 1A and 6D, respectively. For protein content two main QTLs were identified on chromosomes 1B and 6A, respectively. For strength of the dough, three consistent QTLs were detected: two at the same location as those for hardness, on chromosomes 1A and 5D; the third one on chromosome 3B.

3.4. Barley

In barley (*Hordeum vulgare* L.), the first molecular marker linkage map (chromosome 6) was generated on the basis of RFLPs by Kleinhofs *et al.* (1988). To date, more than 1,000 molecular markers, predominantly RFLPs, have been mapped on barley genome. Recently, the genetic linkage maps of four DH populations: Proctor/Nudinka (Heun *et al.*, 1991), Igri/Franka (Graner *et al.*, 1991), Steptoe/Morex (Kleinhofs *et al.*, 1993) and Harrington/TR306 (Kasha and Kleinhofs, 1994), have been integrated into one composite map comprising 898 marker loci covering 1,060 cM, to reduce many large gaps present in each single map.

The AFLP procedure has also provided in this plant a convenient and reliable tool to generate markers to facilitate map construction (Becker *et al.*, 1995; Powell *et al.*, 1997; Qi *et al.*, 1997; Waugh *et al.*, 1997). In this context, Qi *et al.* (1997) increased the marker density by adding 563 AFLP markers segregating in a RIL population. This high-density molecular map contains 561 AFLP markers, three morphological markers, one disease resistance gene and a SSR marker that cover 1,062 cM, corresponding to an average of one marker per 1.9 cM.

The construction of genetic maps of the barley genome permitted the mapping of agronomic qualitative and quantitative traits, including the *ym4* virus resistance gene (Graner and Bauer, 1993), the *denso* dwarfing genes (Laurie *et al.*, 1993), the *liguleless* gene (Pratchett and Laurie, 1994), a photoperiod-response gene (Laurie *et al.*, 1994), and QTLs for yield and yield components, malting quality (Hayes and Iyambo, 1994; Han *et al.*, 1995; Kjaer and Jensen, 1996; Larson *et al.*, 1996; Thomas *et al.*, 1995, 1996; Bezant *et al.*, 1997; Powell *et al.*, 1997), yield-determining physiological characters (Yin *et al.*, 1999), disease resistance (Heun, 1992; Chen *et al.*, 1994; Baches *et al.*, 1995, 1996; Qi *et al.*, 1999; Kicherer *et al.*, 2000), and head shattering (Kandemir *et al.*, 2000).

Molecular marker for qualitative traits can increase the speed of backcross introgressions as an effective tool for selection and for positional cloning. Lahaye *et al.* (1998) described the construction of an high resolution linkage map based on RFLP markers at the *Rar1* locus. This locus is necessary for *Mla-12* specific resistance against *Erysiphe graminis*, of sp. *Hordei*. In addition, the previous authors have assessed physical and genetical distances at the *Rar1* locus as a first step towards the map-based isolation of the gene. A strategy based upon AFLP markers for high-efficiency mapping of morphological mutations and DNA probes to linkage groups was developed by Castiglioni *et al.* (1998). In this study the previous authors have placed 511 AFLP markers on the linkage map derived from the cross Proctor/Nudinka, and assigned loci controlling phenotypic traits to linkage groups by AFLP analysis, using F_2 populations consisting of 30-50 mutant plants derived from crosses of the type "mutant"/Proctor and "mutant"/Nudinka. To map DNA probes, 67 different wildtype barley lines were selected to generate F_2 populations by crossing with Proctor and Nudinka. F_2 plants that were polymorphic for a given RFLP fragment were classified into genotypic classes. Linkage of the RFLP polymorphism to one of the 511 AFLP loci was indicated by cosegregation. The use of this strategy was exemplified by the mapping of the mutation *branched-5* to chromosome 2 and of the DNA probes *Bkn2* and *BM-7* to chromosomes 5 and 1, respectively. A local high resolution genetic map and the application of the AFLP technology has permitted to Büschges *et al.* (1997) the isolation of the *Mlo* gene which is a control element of plant pathogen resistance. This study has also shown the feasibility of a chromosome landing approach in largest genomes such as those of most cereals for which a positional cloning approach had been proposed.

3.5. Sorghum

Since the construction of the first sorghum linkage map using heterologous maize probes (Hulbert *et al.*, 1990), many sorghum molecular maps have been developed and published for this crop in the context of comparative mapping (Binelli *et al.*, 1992; Whitkus *et al.*, 1992; Melake-Berhan *et al.*, 1993; Pereira *et al.*, 1994). In 1994, three different groups (Ragab *et al.*, 1994; Chittenden *et al.*, 1994; Xu *et al.*, 1994) developed RFLP maps using mainly sorghum DNA probes. While the map of Chittenden *et al.* (1994) was developed from an interspecific cross combination (*Sorghum bicolor*/*Sorghum propinquum*), all other maps including that of Xu *et al.* (1994), were established using intraspecific cross combinations. More recently, Dufour *et al.* (1997) published a map based on maize, sugarcane, and cereal anchor probes. Only the maps of Chittenden *et al.* (1994), Pereira *et al.* (1994) contain ten linkage groups. In most cases, F_3 mapping populations were used in map development with the exception of Dufour *et al.* (1997) and Peng *et al.* (1999), in which two RIL populations were used for construction of two composite linkage maps.

The Dufour's map was further saturated with the addition of more heterologous probes and AFLP markers (Boivin *et al.*, 1999). The map includes 343 loci for 11 linkage groups spanning 1,352 cM. Similarly, the sorghum map of Xu *et al.* (1994) has been improved with addition of RFLP markers and more than 100 SSR loci (Kong *et al.*, 2000). Although these mapping efforts

ensure the availability of a large number of probes from different sources, lack of correspondence between the linkage groups of different genetic maps prohibited their effective utilisation for sorghum genome mapping. To achieve this objective, a genetic map of 214 loci with a total map distance of 1,200 cM was constructed by Subudhi and Nguyen (2000), using 98 F₂ sorghum RILs from a cross between the inbred lines B35 and Tx7000. Few sorghum and maize cDNA clones related to photosynthesis and drought stress were positioned on this map for the first time. In addition, five major RFLP maps independently developed in this species were aligned to compare the order of mapped markers. In general, consistent linear order among markers was maintained in all linkage groups. The alignment of these maps has increased the number of markers available for any region of the sorghum genome with many potential applications ranging from fine mapping and MAS to map-based cloning for the improvement of sorghum and related species.

A *S. bicolor* integrated map consisting of 147 SSR loci, and 323 RFLP loci was recently constructed by Bhattaramakki *et al.* (2000b). Most of the SSR primer sets were developed from clones isolated from two sorghum bacterial-artificial-chromosome (BAC) libraries and three enriched sorghum genomic-DNA (gDNA) libraries. The SSR loci are distributed relatively evenly throughout approximately 75% of the 1,406 cM linkage map, but segments of five linkage groups, comprising about 25% of the map, either lack or contain few SSR loci. Mapping of SSR loci isolated from BAC clones located in these areas is likely to be the most efficient method for placing SSR loci in the regions.

In this plant, molecular markers, namely RFLPs and also AFLPs, have been used to map and analyse QTLs associated with agronomically important traits including grain quality, productivity, morphological and agronomical traits (Lin *et al.*, 1995; Pereira and Lee, 1995; Pereira *et al.*, 1995; Tuinstra *et al.*, 1996; Rami *et al.*, 1998), pre-harvest-sprouting resistance (Lijavetzky *et al.*, 2000). Xu *et al.* (2000) have mapped QTLs that control the stay-green, an important post-flowering stress resistance trait in sorghum, and chlorophyll content. They identified four stay-green QTLs, located on three linkage groups. Two QTLs, *Stg1* and *Stg2*, are on linkage group A, with the other two, *Stg3* and *Stg4*, on linkage groups D and J, respectively. *Stg1* and *Stg2*, explaining 13-20% and 20-30% of the phenotypic variability, respectively. They were consistently identified in all trials at different locations in two years. The same authors have also identified three QTLs for chlorophyll content, explaining 25-30% of the phenotypic variability of the progenies tested under post-flowering drought stress. All coincided with the three stay-green QTL regions (*Stg1*, *Stg2*, and *Stg3*) accounting for 46% of the phenotypic variation. The *Stg1* and *Stg2* regions also contain the genes for key photosynthetic enzymes, heat shock proteins, and an abscisic acid responsive gene. Such spatial arrangement shows that linkage group A is important for drought- and heat-stress tolerance and yield production in sorghum. High-resolution mapping and cloning of the consistent stay-green QTLs may help to develop drought-resistant hybrids and to understand the mechanism of drought-induced senescence in plants.

3.6. Other cereals

There are several maps available covering most of the seven rye chromosomes, as reported by Devos *et al.* (1993). Philipp *et al.* (1994), Loarce *et al.* (1996), Senft and Wricke (1996) and Korzun *et al.* (1998), whereas other authors used only selected RFLP markers for tagging genes affecting traits of interest. In that case only single chromosomes or chromosome regions were mapped (Plashke *et al.*, 1993, 1995; Korzun *et al.*, 1996, 1997; Voylovkov *et al.*, 1998).

Comparisons among regions of chromosomes mapped with common markers in different populations indicate that, in most cases, the marker order is identical, although distances are different. Börner and Korzun (1998) have reported a consensus linkage map. This map was

constructed for all seven rye chromosomes using 12 basic RFLP maps. The maps presented contain a total of 413 markers. The number of markers per chromosome varies from 41 (chromosome 3R) to 83 (chromosome 1R). In addition to 374 RFLP and 24 isozyme markers, 15 gene loci were incorporated, determining the traits reduced plant height, self fertility, male sterility restoration, vernalization response, resistance against powdery mildew, chlorophyll deficiency, hairy leaf sheath, hairy peduncle, waxy endosperm, waxless plant, and absence of ligules. The maps presented allow the selection of markers for the fine mapping of certain regions of the rye genome. Since chromosomal rearrangements within the *Triticeae* are known, their utilization can also be extended for mapping in wheat and barley.

Molecular linkage maps have been constructed for diploid (O'Donoghue *et al.*, 1992; Raypati *et al.*, 1994) and hexaploid (O'Donoghue *et al.*, 1995) oat (*Avena sativa* L.). More recently, Portyanko *et al.* (2000) have established an oat linkage map in an $F_{6,7}$ RIL population from the cross Ogle/TAM-0-301 based on the "Cornell Anchor Set" of DNA clones and a set of probes from the John Innes Centre (Norwich, UK) previously mapped in other grass genera. A total of 444 loci were mapped using primarily RFLPs, and secondarily RAPD, SSR, AFLP, intron fragment length polymorphism loci, putative resistance genes, isozymes and morphological traits, resulting in 38 linkage groups. Total map length was 2,042 cM, which is larger than the previously published Kanota/Ogle oat map (O'Donoghue *et al.*, 1995). Approximately, 35, 44, and 28% of the map length is flanked by markers in common with the Kanota/Ogle map, and with diploid oat maps constructed at Cornell University (VanDeynze *et al.*, 1995b), and Iowa State University (Kremer *et al.*, 2000), respectively. Alignment of the hexaploid maps has demonstrated almost complete identity of marker order on both segmental and chromosomal level. Conservation of diploid genome blocks was 92-98%, while chromosome synteny was less because of rearrangements. Multiple marker loci detected sets of homoeological linkage groups, probably corresponding to homoeological chromosome groups, although rearrangements between non-homoeological oat chromosomes were observed.

Similarly, for hexaploid oat a combined RFLP and AFLP map was constructed by Jin *et al.* (2000). The segregation of AFLP markers was scored in two hexaploid oat RIL populations, the Kanota/Ogle RFLP population, and a population derived from Clintland 64 and IL86-5698, barley yellow dwarf virus (BYD) sensitive and BYDV-tolerant lines, respectively. More than 300 AFLP markers were scored in each population, 97 of which were scored in both populations. AFLP markers were linked to RFLP markers in 32 of 36 Kanota/Ogle RFLP linkage groups. The addition of the AFLP markers to the Kanota/Ogle RFLP data set combined markers from four pairs of linkage groups and increased the size of the map from 1,402 cM to 2,351 cM. Thirty linkage groups were observed in the Clintland 64/IL86-5698 population, two of which could be consolidated by comparing the map from both populations. The AFLP and RFLP markers showed very similar distributions in the Kanota/Ogle population, with a tendency of each type of marker to cluster with markers of the same type. The placement of a set of AFLP markers on the Kanota/Ogle linkage map enriched the RFLP map and allowed to relate AFLP markers of agronomically important genes to the reference Kanota/Ogle linkage map.

The hexaploid oat RFLP map has been used to identify QTLs for agronomic traits (Siripoonwiwat *et al.*, 1996), crown rust resistance (Bush and Wise, 1996), and vernalisation responses (Holland *et al.*, 1997). In addition, the significant advances obtained recently in comparative analysis of grass genomes (Gale and Devos, 1998) may make it possible to take advantage of developments in other large scale genomics programmes such as those of rice (Sasaki, 1998), maize (Coe, 1998), and wheat for the study of grasses with relatively sparse linkage maps or large polyploid genomes.

An RFLP-based map consisting of 160 loci was constructed by Wang *et al.* (1998) in one intervarietal cross of foxtail millet [*Setaria italica* (L.) P. Beauv.], Longgu 25/Pagoda Flower

Green. The map comprises nine linkage groups, which were aligned with the nine foxtail millet chromosomes using trisomic lines, and spans 964 cM. The intraspecific map was compared to an interspecific map, constructed in a *S. italica*/*S. viridis* cross. Both the order of the markers and the genetic distances between loci were highly conserved. The segregation data indicate that chromosome 8 in the Longgu 25/Pagoda Flower Green cross carries a gene that strongly affects gamete fertility.

4. PHYSICAL MAPS AND ISOLATION OF AGRONOMICALLY IMPORTANT GENES

Recent development of large DNA fragment (>100kb) manipulation and cloning in yeast (yeast artificial chromosome or YAC; Burke *et al.*, 1987), and bacteria (bacterial artificial chromosome or BAC; Shizuya *et al.*, 1987), has provided powerful tools for generating molecular physical maps of complex genomes, including cereals. Physical maps provide virtually unlimited numbers of DNA markers from any chromosomal region, for gene tagging and mani-

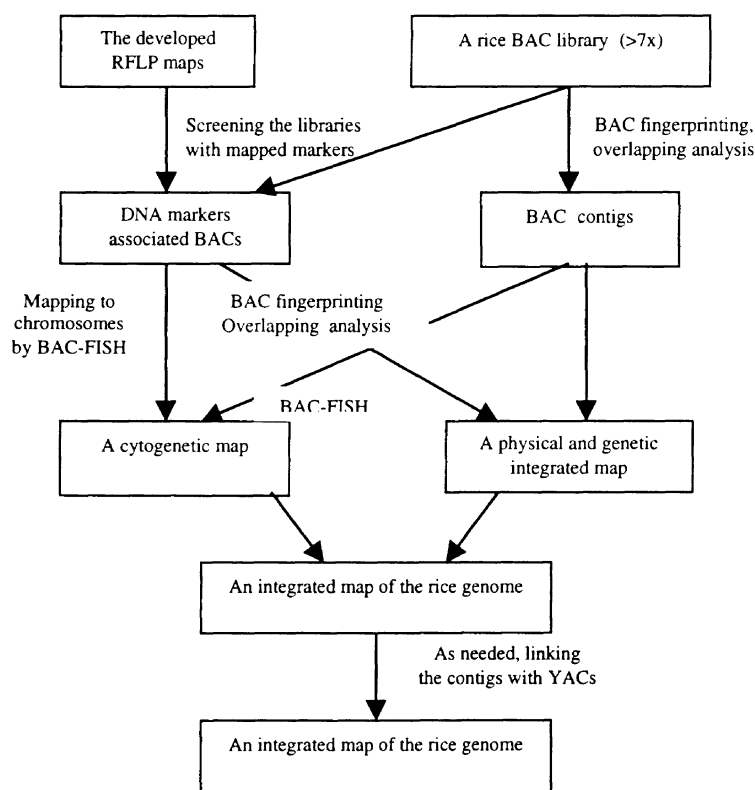


Figure 2. A strategy for physical mapping of the rice genome with BACs (adapted from Zhang and Wing, 1997)

pulation. They also provide an on-line framework for studies in genome molecular structure, organisation and evolution, gene regulation, and gene interaction. The identification, isolation, characterisation, and manipulation of genes will become far more feasible than ever before. Physical maps, therefore, will become central to every type of genetic and molecular inquiry and manipulation, including genome analysis, gene cloning, and crop genetic improvement. YAC libraries have been constructed in rice (Umehara *et al.*, 1995; Tanoue *et al.*, 1997; Wu *et al.*, 1998) and maize (Springer *et al.*, 1994; Faller *et al.*, 2000) and BAC libraries in sorghum (Klein *et al.*, 2000), and rice (Zhang and Wing, 1997). A strategy for physical mapping of the rice genome with BACs is shown in Figure 2 (Zhang and Wing, 1997).

Hong *et al.* (1997) reported that 92% of the rice genome is covered by 631 contigs of various length, which were generated by fingerprinting from a representative and genetically stable BAC library of the cv. Guang Lu Ai 4 genome, with an average insert length of 120 kb. To form the contig map, 565 molecular markers (RFLP, SSR, cDNA and anchor sets) derived from two *O. sativa* varieties were mapped by colony hybridization to the contigs, which were then assigned to and ordered along the particular chromosomes according to marker colinearity. Because of the highly conserved nature of DNA sequences shared among the genomes of rice, barley, wheat, oat, maize, sorghum and sugarcane, 89 of the mapped anchor markers may help to identify rice gene through the information provided by the maps of other cereal genomes, and *vice versa*. Physical distances were determined for hundreds of pairs of adjacent markers, which will facilitate the identification of the rice genes of interest by map-based cloning. The accuracy of clone overlaps in contigs was further confirmed exploiting the existence in contigs of well fitted stacks of marker-lodged clones independently identified by hybridisation. Large scale DNA sequencing of individual chromosomes could now be initiated simply by selecting and sequencing minimally overlapping BAC clones from the contigs.

High density genetic maps coupled with the development of BAC and YAC libraries have been important discoveries leading to the isolation of agronomically important genes. Some examples include isolation of genes (*Xa1*, *Xa21*) for bacterial blight resistance (Sons *et al.*, 1995; Yoshimura *et al.*, 1998) and *Pib* gene for blight resistance in rice (Wang *et al.*, 1999). Sanchez *et al.* (1999) identified BAC contigs flanking the *Xa13* locus for bacterial blight resistance in rice.

Klein *et al.* (2000) have reported the construction of an integrated genetic and physical map of the sorghum genome (750 Mbp). To accomplish this task, these authors have developed a new high-throughput PCR-based method for building BAC contigs and locating BAC clones on the sorghum genetic map. The strategy involved pooling 24,576 sorghum BAC clones (approximately 4x genome equivalents) in six different matrices to create 184 pools of BAC DNAs. DNA fragments from each pool were amplified using AFLP technology. Data from 32 different AFLP primer combinations identified approximately 2,400 BACs and ordered approximately 700 BAC contigs. Analysis of a sorghum RIL mapping population using the same primer pairs located approximately 200 of the BAC contigs on the sorghum genetic map. Analysis of the fingerprint data allowed the identification of 3,366 contigs each containing an average of 5 BACs. Similarly, Faller *et al.* (2000) have made significant progress towards the construction of a whole-genome physical map of maize using a fluorescent fingerprinting approach of a 10x corn BAC library with an average insert size of 155 kb. It was found that gene density per cM is nearly the same between maize and *Arabidopsis*. The authors concluded that positional gene cloning will not be more difficult in maize than in other cereals, once the positional information for all genes is available.

Since many agronomically important genes are known only by their phenotypes, map-based cloning has become an efficient and widely used strategy in isolating such genes mainly in *Arabidopsis* (Lukowitz *et al.*, 2000). Map-based cloning is time-consuming and difficult, it

requires many technical steps and its success is inversely related to the size and complexity of the genome. Therefore the availability of dense physical maps will provide a powerful, and rapid means to isolate numerous economically or biologically important genes.

5. CONCLUSIONS

Molecular mapping of cereal genomes has advanced dramatically in the past 10 years. Currently molecular linkage maps are available for several cereals. As described in this chapter, these maps and their associate technology have been used successfully for a number of applications in plant breeding and genetics. In genome projects high density genetic maps are also central to localising loci of interest in the germplasm and to top-down anchoring physical maps. Altogether these maps are expected to become a key tool in the design and application of new breeding strategies for the improvement of cereal species. It is also expected that the progressive integration of structural and functional genomics with trait development technologies will further accelerate these advances.

6. REFERENCES

- Ahn, S., Bollich, C.N., and Tanksley, S.D. (1992) RFLP tagging of a gene for aroma in rice, *Theor. Appl. Genet.* **84**, 825-828.
- Ahn, S., and Tanksley, S.D. (1993) Comparative linkage maps of the rice and maize genomes, *Proc. Natl. Acad. Sci. USA* **90**, 7980-7984.
- Ahn, S., Anderson, J.A., Sorrells, M.E., and Tanksley, S.D. (1993) Homoeologous relationships of rice, wheat and maize chromosomes, *Mol. Gen. Genet.* **241**, 483-490.
- Ajmone-Marsan, P., Monfredini, G., Ludwig, W.F., Melchinger, A.E., Franceschini, P., Pagnotto, G., and Motto, M. (1995) In an elite cross of maize a major quantitative trait locus controls one-fourth of the genetic variation for grain yield, *Theor. Appl. Genet.* **90**, 415-424.
- Ajmone-Marsan, P., Gorni, C., Chitto, A., Redaelli, R., van Vijk, R., Stam, P., and Motto, M. (2001) Identification of QTLs for grain yield and grain-related traits of maize (*Zea mays* L.) using an AFLP map, different testers, and cofactor analysis, *Theor. Appl. Genet.* **102**, 230-243.
- Ali, M.L., Pathan, M.S., Zhang, J., Bai, G., Sarkarung, S., and Nguyen, H.T. (2000) Mapping QTLs for root traits in a recombinant inbred population from two *indica* ecotypes in rice, *Theor. Appl. Genet.* **101**, 756-766.
- Anderson, J.A., Aihara, Y., Sorrells, M.E., and Tanksley, S.D. (1992) Development of a chromosomal arm map for wheat based on RFLP markers, *Theor. Appl. Genet.* **83**, 1035-1043.
- Anderson, J.A., Sorrells, M.E., and Tanksley, S.D. (1993) RFLP analysis of genomic regions associated with resistance to preharvest sprouting in wheat, *Crop Sci.* **33**, 453-459.
- Armstrong, C.L., Romero-Severson, J., and Hodge, T.K. (1992) Improved tissue culture response of an elite maize inbred through backcross breeding, and identification of chromosomal regions important for regeneration, by RFLP analysis, *Theor. Appl. Genet.* **84**, 755-762.
- Austin, D.F., and Lee, M. (1996a) Genetic resolution and verification of quantitative trait loci for flowering and plant height with recombinant inbred lines of maize, *Genome* **39**, 957-968.
- Austin, D.F., and Lee, M. (1996b) Comparative mapping in F_2 and F_6 generations of quantitative trait loci for grain yield and yield components in maize, *Theor. Appl. Genet.* **2**, 817-826.
- Autrique, E., Singh, R.P., Tanksley, S.D., and Sorrells, M.E. (1995) Molecular markers for four leaf rust resistance genes introgressed into wheat from wild relatives, *Genome* **38**, 75-83.
- Backes, G., Graner, A., Foroughi-Wehr, B., Foschbeck, G., Wenzel, G., and Jahoor, A. (1995) Localization of quantitative trait loci (QTL) for agronomic important characters by the use of a RFLP map in barley (*Hordeum vulgare* L.), *Theor. Appl. Genet.* **90**, 294-302.
- Backes, G., Schwarz, G., Wenzel, G., and Jahoor, A. (1996) Comparison between QTL analysis on powdery mildew resistance in barley based on detached primary leaves and on field data, *Plant Breed.* **115**, 419-421.
- Beavis, B. (1994) The power and deceit of QTL experiments: Lessons from comparative QTL studies, in *Proc. of the Forty-Ninth Annual Corn and Sorghum Industry Res. Conf.*, ASTA, Washington, DC, pp. 250-266.
- Beavis, W.D., Lee, M., Grant, D., Hallauer, A.R., Owens, T., and *et. al.* (1992) The influence of random mating on recombination among RFLP loci, *Maize Gen. Coop. Newsletters* **66**, 52-53.
- Beavis, W.D., Smith, O.S., Grant, D., and Fincher, R. (1994) Identification of quantitative trait loci using a small sample of topcrossed and F4 progeny from maize, *Crop Sci.* **34**, 882-896.

- Becker, J., Vos, P., Kuiper, M., Salamini, F., and Heun, M. (1995) Combined mapping of AFLP and RFLP markers in barley, *Mol. Gen. Genet.* **249**, 65-73.
- Beckett, J.B. (1978) B-A translocations in maize. I. Use in locating genes by chromosome arms, *J. Hered.* **69**, 27-36.
- Berke, T.G., and Rocheford, T.R. (1999) Quantitative trait loci for tassel traits in maize, *Crop Sci.* **39**, 1439-1443.
- Bezant, J., Laurie, D., Pratchett, N., Chojek, J., and Kearey, M. (1997) Mapping QTLs controlling yield and yield components in a spring barley (*Hordeum vulgare* L.) cross using marker regression, *Mol. Breed.* **3**, 29-38.
- Bhatramakki, D., Ching, A., Morgante, M., Dolan, M., Register, J., Smith, H., Tingey, S., and Rafalski, A. (2000a) Single nucleotide polymorphisms (SNPs) in the 3'-untranslated flanks of maize genes: identification, genetic mapping and haplotype structure, in *New horizons in marker technologies and their application for cereal genomics*, Sigma Meeting, Bertinoro, Italy, April 27-29, 2000.
- Bhatramakki, D., Dong, J., Chhabra, A.K., and Hart, G.E. (2000b) An integrated SSR and RFLP linkage map of *Sorghum bicolor* (L.) Moench, *Genome* (in press).
- Binelli, G., Gianfranceschi, L., Pè, M.E., Taramino, G., Busso, C., Stenhouse, J., and Ottaviano, E. (1992) Similarity of maize and sorghum genomes as revealed by maize RFLP probes, *Theor. Appl. Genet.* **84**, 10-16.
- Blanco, A., Bellomo, M.P., Lotti, C., Maniglio, T., Pasqualone, A., Simeone, R., Troccoli, A., and Di Fonzo, N. (1998) Genetic mapping of sedimentation volume across environments using recombinant inbred lines of durum wheat, *Plant Breed.* **117**, 413-417.
- Bohn, M., Khairallah, M.M., Gonzalez-de-Leon, D., Hoisington, D., Utz, H.F., Deutsch, J.A., Jewell, D.C., Mihm, J.A., and Melchinger, A.E. (1996) QTL mapping in tropical maize: I. Genomic regions affecting leaf feeding resistance to sugarcane borer and other traits, *Crop Sci.* **36**, 1352-1361.
- Bohn, M., Khairallah, M.M., Jiang, C., González-de-León, D., Hoisington, D.A., Utz, H.F., Deutsch, J.A., Jewell, D.C., Mihm, J.A., and Melchinger, A.E. (1997) QTL mapping in Tropical Maize: II. Comparison of Genomic Regions for Resistance to *Diatraea* spp., *Crop Sci.* **37**, 1892-1902.
- Boivin, K., Deu, M., Rami, J.-F., Trouche, G., and Hamon, P. (1999) Towards a saturated sorghum map using RFLP and AFLP markers, *Theor. Appl. Genet.* **98**, 320-328.
- Börner, A., and Korzum, V. (1998) A consensus linkage map of rye (*Secale cereale* L.) including 374 RFLPs 24 isozymes and 15 gene loci, *Theor. Appl. Genet.* **97**, 1279-1288.
- Botstein, D., White, R.L., Skolnick, M., and Davis R.W. (1980) Construction of a genetic linkage map using restriction fragment length polymorphisms, *Ann. J. Human Genet.* **32**, 314-331.
- Burke, D.T., Carle, G.F., and Olson, M.V. (1987) Cloning of large segments of exogenous DNA into yeast by means of artificial chromosome vectors, *Science* **236**, 805-811.
- Burr, B.M., Burr, F.A., Thompson, K.H., Albertson, M., and Stuber, C.W. (1988) Gene mapping with recombinant inbreds in maize, *Genetics* **118**, 519-526.
- Büsches, R., Hollricher, K., Panstruga, R., Simons, G., Wolter, M., Frijters, A., and *et al.* (1997) The barley *Mlo* gene: a novel control element of plant pathogen resistance, *Cell* **88**, 695-705.
- Bush, A.L., and Wise, R.P. (1996) Crown rust resistance loci on linkage group 4 and 13 in cultivated oat, *J. Hered.* **87**, 427-432.
- Byrne, P.F., McMullen, M.D., Snook, M.E., Musket, T., Theuri, J.M., and *et al.* (1996) Quantitative trait loci and metabolic pathways: genetic control of the concentration of maysin, a corn earworm resistance factor, in maize silks, *Proc. Natl. Acad. Sci. USA* **93**, 8820-8825.
- Cadalen, T., Boeuf, C., Bernard, S., and Bernard, M. (1997) An intervarietal molecular map in *Triticum aestivum* L. em. Thell and comparison with a map from a wide cross, *Theor. Appl. Genet.* **94**, 367-377.
- Cai, H.W., and Morishima, H. (2000) Genomic regions affecting seed shattering and seed dormancy in rice, *Theor. Appl. Genet.* **100**, 840-846.
- Carlson, W.R., (1988) The cytogenetics of corn, in G.F. Sprague and J.W. Dudley (eds.), *Corn and Corn Improvement*, American Society of Agronomy, Madison, pp. 259-344.
- Castiglioni, P., Pozzi, C., Heun, M., Terzi, V., Muller, K.J., Rohde, W., and Salamini, F. (1998) An AFLP-based procedure for the efficient mapping of mutations and DNA probes in barley, *Genetics* **149**, 2039-2056.
- Castiglioni, P., Ajmone-Marsan, P., van Wijk, R., and Motto, M. (1999) AFLP[®] markers in a molecular linkage map of maize: codominant scoring and linkage group distribution, *Theor. Appl. Genet.* **99**, 425-431.
- Causse, M., Fulton, T.M., Cho, Y.G., Ahn, S.N., Chunwongse, E.J., Wu, K., Xiao, J., Yu, Z., and *et al.* (1994) Saturated molecular map of the rice genome based on an interspecific backcross population, *Genomics* **138**, 1251-1274.
- Causse, M., Santoni, S., Damerval, C., Maurice, A., Charcosset, A., Deatrick, J., and de Vienne, D. (1996) A composite map of expressed sequences in maize, *Genome* **39**, 418-432.
- Chantret, N., Sourdille, P., Röder, M., Tavaud, M., Bernard, M., and Doussinault, G. (2000) Location and mapping of the powdery mildew resistance gene MIRE and detection of a resistance QTL by bulked segregant analysis (BSA) with microsatellites in wheat, *Theor. Appl. Genet.* **100**, 1217-1224.
- Chao, S., Sharp, P.J., Worland, A.J., Warham, E.J., Koebner, R.M.D., and Gale, M.D. (1989) RFLP-based genetic maps of wheat homoeologous group 7 chromosomes, *Theor. Appl. Genet.* **78**, 495-504.

- Chen, F.Q., Prehn, D., Hayes, P.M., Mulrooney, D., Corey, A., Vivar, H. (1994) Mapping genes for resistance to barley stripe rust (*Puccinia striiformis* f. sp. *Hordei*), *Theor. Appl. Genet.* **88**, 215-219.
- Chittenden, L.M., Schertz, K.F., Lin, V.-R., Wing, R. A., and Paterson, A.H. (1994) A detailed RFLP map of *Sorghum bicolor* x *S. propinquum*, suitable for high-density mapping, suggests ancestral duplication of *Sorghum* chromosomes or chromosomal segments, *Theor. Appl. Genet.* **87**, 925-933.
- Cho, Y.G., Eun, M.Y., McCouch, S.R., and Chae, Y.A. (1994) Molecular mapping and genotypic selection for the semi-dwarf gene, *sd-1* in rice (*Oryza sativa* L.), *Theor. Appl. Genet.* **89**, 54-59.
- Cho, Y.G., McCouch, S.R., Kuiper, M., Kang, M.R., Pot, J., Groenen, J.T.M., and Eun, M.Y. (1998) Integrated map of AFLP, SSLP, and RFLP markers using a recombinant inbred population of rice (*Oryza sativa* L.), *Theor. Appl. Genet.* **97**, 370-380.
- Coe, E.H., Hoisington, D.A., and Neuffer, M.G. (1987) Linkage map of corn (maize) (*Zea mays* L.), *Maize Genet. Coop. Newsletters* **61**, 116-147.
- Coe, E., Hancock, D., Kowaleski, S., and Polacco, M. (1995) Gene list and working maps, *Maize Genet. Coop. Newsletters* **69**, 191-267.
- Coe, E.H. (1998) Potentials of the national core genome initiative, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 2029-2032.
- Davis, G., McMullen, M.D., Baysdofer, C., Musket, T., Grant, D., Staebell, M., Xu, G., Polacco, M., Koster, L., Melia-Hancock, S., Houchins, K., Chao, S., and Coe, E.H. (1999) A maize map standard with sequenced core markers, grass genome reference points and 932 expressed sequences tagged sites (ESTs) in a 1736-locus map, *Genetics* **152**, 1137-1172.
- Dedryver, F., Jubier, M.F., Thouvenin, J., and Goyeau, H. (1996) Molecular markers linked to the leaf rust resistance gene *Lr24* in different wheat cultivars, *Genome* **39**, 830-835.
- Devey, M.E., and Hart, G.E. (1993) Chromosomal localization of intergenomic RFLP loci in hexaploid wheat, *Genome* **36**, 913-918.
- Devos, K.M., Atkinson, M.D., Chinoy, C.N., Liu, C.J., and Gale, M.D. (1992) RFLP-based genetic map of the homoeologous group 3 chromosomes of wheat and rye, *Theor. Appl. Genet.* **83**, 931-939.
- Devos, K.M., Millan, T., and Gale, M.D. (1993) Comparative RFLP maps of homoeologous group 2 chromosomes of wheat, rye and barley, *Theor. Appl. Genet.* **85**, 784-792.
- Devos, K.M., Chao, S., Qy, L., Simonetti, M.C., and Gale, M. (1994) Relationships between chromosome 9 of maize and wheat homoeologous group 7 chromosomes, *Genetics* **138**, 1287-1292.
- Doebley, J., and Stec, A. (1991) Genetic analysis of the morphological differences between maize and teosinte, *Genetics* **129**, 285-295.
- Dong, N.V., Subudhi, P.K., Luong, P.N., Quang, V.D., Quy, T.D., Zheng, H.G., Wang, B., and Nguyen, H.T. (2000) Molecular mapping of a rice gene conditioning thermosensitive genic male sterility using AFLP, RFLP and SSR techniques, *Theor. Appl. Genet.* **100**, 727-734.
- Dubcovsky, J., Luo, M.C., Zhong, G.Y., Bransteiter, R., Desai, A., et al. (1996) Genetic map of diploid wheat, *Triticum monococcum* L., and its comparison with maps of *Hordeum vulgare* L., *Genetics* **143**, 983-999.
- Dubcovsky, J., Lukaszewski, A.J., Echaide, M., Antonelli, E.F. and Porter, D.R. (1998) Molecular characterization of two *Triticum speltoides* interstitial translocations carrying leaf rust and greenbug resistance genes, *Crop Sci.* **38**, 1655-1660.
- Dufour, R., Deu, M., Grivet, L., D'Hont, A., Paulet, F., Bouet, A., Lanaud, C., Glaszmann, J.C., and Hamon, P. (1997) Construction of a composite sorghum genome map and comparison with sugarcane, a related complex polyploid, *Theor. Appl. Genet.* **94**, 409-418.
- Edwards, M.D., Stuber, C.W., and Wendel, J.F. (1987) Molecular-marker facilitated investigations of quantitative trait loci in maize. I. Numbers, genomic distributions and types of gene action, *Genetics* **116**, 113-125.
- Edwards, M.D., Helentjaris, T., Wright, S., and Stuber, C.W. (1992) Molecular-marker-facilitates investigations of quantitative trait loci in maize. 4. Analysis based on genome saturation with isozyme and restriction fragment length polymorphism markers, *Theor. Appl. Genet.* **83**, 765-774.
- Evola, S.V., Burr, F.A., and Burr, B. (1986) The suitability of restriction fragment length polymorphisms as genetic markers in maize, *Theor. Appl. Genet.* **71**, 765-771.
- Faller, M.L., Fenger, K.A., Meyers, B.C., Dolan, M., Tingey, S.V., and Morgante, M. (2000) Genome sequencing and biology integrating genetic and physical maps for positional cloning: ESTs, SNPs and Bacs, in *New horizons in marker technologies and their application for cereal genomics*, Sigma Meeting, Bertinoro, Italy, April 27-29, 2000.
- Frei, O.M., Stuber, C.W., and Goodman, M.M. (1986) Uses of allozymes as genetic markers for predicting performance in maize single-cross hybrids, *Crop Sci.* **26**, 37-42.
- Freymark, P.J., Lee, M., Woodman, W.L., and Martinson, C.A. (1993) Quantitative and qualitative trait loci affecting host-plant response to *Exserohilum turcicum* in maize (*Zea mays* L.), *Theor. Appl. Genet.* **87**, 537-544.
- Frova, C., and Sari-Gorla, M. (1993) Quantitative expression of maize HSPs: genetic dissection and association with thermotolerance, *Theor. Appl. Genet.* **86**, 213-220.
- Gale, M.D., and Devos, K.M. (1998) Plant comparative genetics after 10 years, *Science* **282**, 656-659.

- Gardiner, J.M., Coe, E.H., Melia-Hancock, S., Hoisington, D.A., and Chao, S. (1993) Development of a core RFLP map in maize using and immortalized F2 population, *Genetics* **134**, 917-930.
- Gill, K.S., Lubbers, E.L., Gill, B.S., Raupp, W.J., and Cox, T.S. (1991) A genetic linkage map of *Triticum tauschii* (DD) and its relationship to the D genome of bread wheat (AABBDD), *Genome* **34**, 362-374.
- Goldman, I.L., Rocheford, T.R., and Dudley, J.W. (1993) Quantitative trait loci influencing protein and starch concentration in the Illinois Long Term Selection maize strains, *Theor. Appl. Genet.* **87**, 217-224.
- Goldman, I.L., Rocheford, T.R., and Dudley, J.W. (1994) Molecular markers associated with maize kernel oil concentration in an Illinois high protein x Illinois low protein cross, *Crop Sci.* **34**, 908-919.
- Graham, G.I., Wolff, D.W., and Stuber, W. (1995) Characterization of a yield quantitative trait locus on chromosome five of maize by fine mapping, *Crop Sci.* **37**, 1601-1610.
- Graner, A., Jahoor, A., Schondelmaier, J., Siedler, H., Pillen, K., Fischbeck, G., Wenzel, G., and Herrmann, R.G. (1991) Construction of an RFLP map of barley, *Theor. Appl. Genet.* **83**, 250-256.
- Graner, A., and Bauer, E. (1993) RFLP mapping of the *ym4* virus resistance gene in barley, *Theor. Appl. Genet.* **86**, 689-693.
- Han, F., Ullrich, S.E., Chirat, S., Menteur, S., Jestin, L., Sarrafi, A., Hayes, P.M., and *et al.* (1995) Mapping of β -glucanase activity loci in barley grain and malt, *Theor. Appl. Genet.* **91**, 921-927.
- Hart, G.E., Gale, M.D., and McIntosh, R.A. (1993) Linkage maps of *Triticum aestivum* (hexaploid wheat, $2n = 42$, genomes A, B & D) and *T. tauschii* ($2n = 14$, genome D.), in D. Hoisington and A. McNa (eds.), *Progress in genome mapping of wheat and related species*, *Proc. of the 3rd Public Workshop of the International Triticeae Mapping Initiative*, Mexico, D.F. Sept. 22-26 (1992), *Cimmyt*, pp. 32-46.
- Harushima, Y., Yano, M., Shomura, A., Sato, M., Shimano, T., and *et al.* (1998) A high-density rice genetic linkage map with 2275 markers using a single F2 population, *Genetics* **148**, 479-494.
- Hayes, P.M., and Iyambo, O. (1994) The North American barley genome mapping project. Summary of QTL effects in the Steptoe x Morex population, *Barley Genet. Newsletters* **23**, 98-133.
- Helentjaris, T., Slocum, M., Wright, S., Schaefer, A., and Nienhuis, J. (1986) Construction of genetic linkage maps in maize and tomato using restriction fragment length polymorphism. *Theor. Appl. Genet.* **72**, 761-769.
- Helguera, M., Khan, I.A., and Dubcovsky, J. (2000) Development of PCR markers for the wheat leaf rust resistance gene *Lr47*, *Theor. Appl. Genet.* **100**, 1137-1143.
- Heun, M., Kennedy, A.E., Anderson, J.A., Lapitan, B.L.V., Sorrells, M.E., and Tanksley, S.D. (1991) Construction of a restriction fragment length polymorphism map for barley (*Hordeum vulgare*), *Genome* **34**, 437-447.
- Heun, M. (1992) Mapping quantitative powdery mildew resistance of barley using a restriction fragment length polymorphism map, *Genome* **35**, 1019-1025.
- Holland, J.B., Moser, H.S., O'Donoghue, L.S., and Lee, M. (1997) QTLs and epistasis associated with vernalization response in oat, *Crop Sci.* **37**, 1306-1316.
- Holland, J.B., Uhr, D.V., Jeffers, D., and Goodman, M.M. (1998) Inheritance of resistance to southern corn rust in tropical-by-corn-belt maize populations, *Theor. Appl. Genet.* **96**, 232-241.
- Hong, G., Qian, Y., Yu, S., Hu, X., Zhu, J., Tao, W., Li, W., Su, C., Zhao, H., Qiu, L., Yu, D., Liu, X., Wu, B., Zhang, X., and Zhao, W. (1997) A 120 kilobase resolution contig map of the rice genome, *DNA Seq.* **7**, 319-335.
- Huang, N., McCouch, S.R., Mew, T., Parco, A., and Guiderdoni, E. (1994) Development of a RFLP map from a doubled haploid population of rice, *Rice Genet. Newslett.* **11**, 134-137.
- Hulbert, S.H., Richter, T.E., Axtell, J.D., and Bennetzen, J.L. (1990) Genetic mapping and characterization of sorghum and related crops by means of maize DNA probes, *Proc. Natl. Acad. Sci. USA* **87**, 4251-4255.
- Ishikawa, R., Morishima, H., Kinoshita, T., Harada, T., Nuzeri, M., and *et al.* (1991) Linkage analysis of nine isozyme genes on the conventional linkage map in rice, *Jpn. J. Breed.* **41**, 265-272.
- Jin, H., Domier, L.L., Shen, X., and Kolb, F.L. (2000) Combined AFLP and RFLP mapping in two hexaploid oat recombinant inbred populations, *Genome* **43**, 94-100.
- Jung, M., Weldekidan, T., Schaff, D., Paterson, A., Tingey, S., and Hawk, J. (1994) Generation-means analysis and quantitative trait locus mapping of anthracnose stalk rot genes in maize, *Theor. Appl. Genet.* **89**, 413-418.
- Kandemir, N., Kudrna, D.A., Ullrich, S.E., and Kleinhofs, A. (2000) Molecular marker assisted genetic analysis of head shattering in six-rowed barley, *Theor. Appl. Genet.* **101**, 203-210.
- Kasha, K.J., and Kleinhofs, A. (1994) Mapping of barley cross Harrington x TR306, *Barley Genet. Newsletters* **23**, 65-69.
- Khavkin, E., and Coe, E. (1997) Mapped genomic locations for developmental functions and QTLs reflect concerted groups in maize (*Zea mays* L.), *Theor. Appl. Genet.* **95**, 343-352.
- Kicherer, S., Backes, G., Walther, U., and Jahoor, A. (2000) Localising QTLs for leaf rust resistance and agronomic traits in barley (*Hordeum vulgare* L.), *Theor. Appl. Genet.* **100**, 881-888.
- Kjær, B., and Jensen, J. (1996) Quantitative trait loci for grain yield and yield components in a cross between a six-rowed and a two-rowed barley, *Euphytica* **90**, 39-48.
- Klein, P.E., Klein, R.R., Cartinhour, S.W., Ulanich, P.E., Dong, J., Obert, J.A., Morishige, D. T., Schlueter, S.D., Childs, K.L., Ale, M., and Mullet, J.E. (2000) A high-throughput AFLP-based method for constructing integrated

- genetic and physical maps: progress toward a sorghum genome map, *Genome Res.* **10**, 789-807.
- Kleinhofs, A., Chao, S., and Sharp, P.J. (1988) Mapping of nitrate reductase genes in barley and wheat, in T.E. Miller and R.M.D. Koebner (eds.) *Proc. 7th International Wheat Genetics Symposium*, Bath Press, Bath, pp. 541-546.
- Kleinhofs, A., Kilian, A., Saghai Maroof, M.A., Biyashev, R.M., Hayes, P.M., and *et al.* (1993) A molecular, isozyme and morphological map of the barley (*Hordeum vulgare*) genome, *Theor. Appl. Genet.* **86**, 705-712.
- Koester, R.P., Sisco, P.H., and Stuber, C.W. (1993) Identification of quantitative trait loci controlling days to flowering and plant height in two near isogenic lines of maize, *Crop Sci.* **33**, 1209-1216.
- Kong, L., Dong, J., and Hart, G.E. (2000) Characteristics, linkage-map positions, and allelic differentiation of *Sorghum bicolor* (L.) Moench DNA simple-sequence repeats (SSRs) (in press).
- Korzun, V., Meiz, G. and Börner, A. (1996) RFLP mapping of the dwarfing (*Ddwl*) and hairy peduncle (*Hp*) genes on chromosome 5 of rye (*Secale cereale* L.), *Theor. Appl. Genet.* **92**, 1073-1077.
- Korzun, V., Malyshev, S., Voylovkov, A. and Börner, A. (1997) RFLP-based mapping of three mutant loci in rye (*Secale cereale* L.) and their relation to homoeologous loci within the *Gramineae*, *Theor. Appl. Genet.* **95**, 468-473.
- Korzun, V., Malyshev, S., Kartel, N., Westermann, T., Weber, W.E., and Börner, A. (1998) A genetic linkage map of rye (*Secale cereale* L.), *Theor. Appl. Genet.* **95**, 468-473.
- Kremer, C.A., Lee, M. and Holland, J.B. (2000) An RFLP-based linkage map of a diploid *Avena* recombinant inbred line population. *Genome* (in press).
- Kurata, N., Moore, G., Nagumara, Y., Foote, T., Yano, M., Minobe, Y., and Gale, M. (1994) Conservation of genome structure between rice and wheat, *Bio/Technology* **12**, 276-278.
- Lagudah, E.S., Appels, R., Brown, A.H.D., and McNeil, D.F. (1991) The molecular-genetic analysis of *Triticum tauschii*, the D genome donor to hexaploid wheat, *Genome* **34**, 375-386.
- Lahaye, T., Hartmann, S., Töpsch, S., Freialdenhoven, A., Yano, M., and Schulze-Lefert, P. (1998) High-resolution genetic and physical mapping of the *Rar1* locus in barley, *Theor. Appl. Genet.* **97**, 526-534.
- Larson, S.R., Kadyrzhanova, D., McDonald, C., Sorrells, M., and Blake, T.K. (1996) Evaluation of barley chromosome-3 yield QTLs in a backcross F2 population using STS-PCR, *Theor. Appl. Genet.* **93**, 618-625.
- Laurie, D.A., Pratchett, N., Romero, C., Simpson, E., and Snape, J.W. (1993) Assignment of the *denso* dwarfing gene to the long arm of chromosome 3 (3H) of barley by use of RFLP markers, *Plant Breed.* **111**, 198-203.
- Laurie, D.A., Pratchett, N., Bezant, J.H., and Snape, J.W. (1994) Genetic analysis of a photoperiod response gene on the short arm of chromosome 2 (2H) of *Hordeum vulgare* (barley), *Heredity* **72**, 619-627.
- Lebreton, C., Lazic-Jancic, V., Steed, A., Pekic, S., and Quarrie, S.A. (1995) Identification of QTL for drought responses in maize and their use in testing causal relationships between traits, *J. Exp. Bot.* **46**, 853-865.
- Lijavetzky, D., Martinez, M.C., Carrari, F., and Hopp, H.E. (2000) QTL analysis and mapping of pre-harvest sprouting resistance in Sorghum, *Euphytica* **112**, 125-135.
- Lin, Y.R., Schertz, K.F., and Paterson, A.H. (1995) Comparative analysis of QTLs affecting plant height and maturity across the *Poaceae*, in reference to an interspecific sorghum population, *Genetics* **140**, 391-411.
- Liu, A., Li, H., Zhang, Q., Jiang, S., and *et al.* (1992) Mapping a wide compatibility gene of rice in relation to RFLP markers, *J. Huazhong Agric. Univ.* **11**, 213-219.
- Liu, Y.G., and Tsunewaki, K. (1991) Restriction fragment length polymorphism analysis of wheat. II. Linkage maps of the RFLP sites in common wheat, *Jpn. J. Genet.* **66**, 617-633.
- Loarce, Y., Hueros, G. and Ferrer, E. (1996) A molecular linkage map of rye, *Theor. Appl. Genet.* **93**, 1112-1118.
- Lübberstedt, T., Melchinger, A.E., Schön, C., Utz, H.F., and Klein, D. (1997) QTL mapping in testcrosses of European Flint lines of maize. I. Comparison of different testers for forage yield traits, *Crop Sci.* **37**, 921-931.
- Lübberstedt, T., Klein, D., and Melchinger, A.E. (1998) Comparative QTL mapping of resistance to *Ustilago maydis* across four populations of European flint-maize, *Theor. Appl. Genet.* **97**, 1321-1330.
- Lübberstedt, T., Xia, X.C., Tan, G., Liu, T., and Melchinger, A.E. (1999) QTL mapping of resistance to *Sporisorium reilianum* in maize, *Theor. Appl. Genet.* **99**, 593-598.
- Lukowitz, W., Gillmor, C.S., and Scheible, W.R. (2000) Positional cloning in Arabidopsis. Why it feels good to have a genome initiative working for you, *Plant Physiol.* **123**, 795-805.
- Ma, Z.Q., and Sorrells, M.E. (1995) Genetic analysis of fertility restoration in wheat using restriction fragment length polymorphisms, *Crop Sci.* **35**, 1137-1143.
- Mackill, D.J., Salam, M.A., Wang, Z.Y., and Tanksley, S.D. (1993) A major photoperiod-sensitivity gene tagged with RFLP and isozyme markers in rice, *Theor. Appl. Genet.* **85**, 536-540.
- Maheswaran, M., Subudhi, P.K., Nandi, S., Xu, J.C., Parco, A., Yang, D.C., and Huang, N. (1997) Polymorphism, distribution, and segregation of AFLP markers in a doubled-haploid rice population, *Theor. Appl. Genet.* **94**, 39-45.
- Marino, C.L., Nelson, J.C., Lu, Y.H., Sorrells, M.E., Leroy, P., and *et al.* (1996) Molecular genetic maps of the group 6 chromosomes of hexaploid wheat (*Triticum aestivum* L. em. Thell), *Genome* **39**, 359-366.
- Martinant, J.P., Cadalen, T., Billot, A., Chartier, S., Leroy, P., and *et al.* (1998) Genetic analysis of water extractable arabinolxilans in bread wheat endosperm, *Theor. Appl. Genet.* **97**, 1069-1075.

- McCouch, S.R., Kochert, G., Yu, Z.H., Wang, Z.Y., Knushi, G.S., and *et al.* (1988) Molecular mapping of rice chromosomes, *Theor. Appl. Genet.* **76**, 815-829.
- McCouch, S.R., and Tanksley, S.D. (1991) Development and use of restriction fragment length polymorphism in rice breeding and genetics, in G. Toenniessen and G. Khush (eds.), *Rice Biotechnology*, CAB International, Tucson, Ariz., pp. 135-155.
- McIntosh, R.A., Wellings, C.R., and Park, R.F. (1995) *Wheat rusts, an atlas of resistance genes*, CSIRO, Melbourne, Australia.
- Melake-Berhan, A., Hulbert, S.H., Butler, L.G., and Bennetzen, J.L. (1993) Structure and evolution of the genomes of *Sorghum bicolor* and *Zea mays*, *Theor. Appl. Genet.* **86**, 598-604.
- Melchinger, A.E., Kuntze, L., Gumber, R.K., Lübberstedt, T., and Fuchs, E. (1998) Genetic basis of resistance to sugarcane mosaic virus in European maize germplasm, *Theor. Appl. Genet.* **96**, 1151-1161.
- Mohan, M., Nair, S., and Bennett, J. (1993) Mapping of a rice gene for resistance to biotype-1 of gall midge (*Orseolia oryzae*) by RFLP and RAPD analyses, in *Proc. of the Sixth Annual Meeting of the International Program on Rice Biotechnology*, Feb. 1-5, Chiang Mai, Thailand, pp. 13.
- Mohan, M., Nair, S., Bhagwat, A., Krishna, T.G., Yano, M., Bhatia, C.R., and Sasaki, T. (1997) Genome mapping, molecular markers and marker-assisted selection in crop plants, *Molec. Breeding* **3**, 87-103.
- Moore, G., Devos, K.M., Wang, Z. and Gale, M.D. (1995) Grasses, line up and form a circle, *Curr. Biology* **5**, 737-739.
- Nagamura, Y., Antonio, B.A., Fukuda, A., Harushima, Y., Inoue, T., and *et al.* (1993) A high density STS and EST linkage map of rice, *Rice Genome* **2**, 3.
- Nelson, J.C., Van Deynze, A.E., Autrique, E., Sorrells, M.E., Lu, Y.H., Merlino, M., Atkinson, M., and Leroy, P. (1995a) Molecular mapping of wheat. Homeologous group 2, *Genome* **38**, 516-524.
- Nelson, J.C., Van Deynze, A.E., Lu, Y.H., Autrique, E., Sorrells, M.E., Negre, S., Bernard, M., and Leroy, P. (1995b) Molecular mapping of wheat. Homeologous group 3, *Genome* **38**, 525-533.
- Nelson, J.C., Sorrells, M.E., Van Deynze, A.E., Lu, Y.H., Atkinson, M., Leroy, P., Faris, J., and Anderson, J. (1995c). Molecular mapping of wheat. Major genes and rearrangements in homoeologous groups 4, 5, and 7, *Genetics* **141**, 721-731.
- Nelson, J.C., Singh, R.P., Autrique, J.E. and Sorrells, M.E. (1997) Mapping genes conferring and suppressing leaf rust resistance in wheat, *Crop Sci.* **37**, 1928-1935.
- Ni, J.J., Wu, P., Senedhira, D., and Huang, N. (1998) Mapping QTLs for phosphorus deficiency tolerance in rice (*Oryza sativa* L.), *Theor. Appl. Genet.* **97**, 1361-1369.
- O'Brien, S.J. (1993) *Genetic maps*, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.
- O'Donoghue, L.S., Wang, Z., Röder, M., Kneen, B., Leggett, M., Sorrells, M.E. and Tanksley, S.D. (1992) An RFLP-based linkage map of oat based on a cross between two diploid taxa (*Avena atlantica* x *A. hirtula*), *Genome* **35**, 765-771.
- O'Donoghue, L.S., Kianian, S.F., Rayapati, P.J., Penner, G.A., Sorrells, M.E., Tanksley, S.D., Phillips, R.L., Rines, H.W., Lee, M., Fedak, G., Molnar, S.J., Hoffman, D., Salas, C.A., Wu, B., Autrique, E., and Van Deynze, A. (1995) A molecular map of cultivated oat, *Genome* **38**, 368-380.
- O'Donoghue, L.S., Chong, J., Wight, C.P., Fedak, G., and Molnar, S.J. (1996) Localization of stem rust resistance genes and associated molecular markers in cultivated oat, *Phytopathology* **86**, 719-727.
- Ottaviano, E., Sari-Gorla, M., Pè, E., and Fropa, C. (1991) Molecular markers (RFLPs and HSPs) for the genetic dissection of thermotolerance in maize, *Theor. Appl. Genet.* **81**, 713-719.
- Pejic, I., Ajmone-Marsan, P., Morgante, M., Kozumplick, V., Castiglioni, P., Taramino, G., and Motto, M. (1998) Comparative analysis of genetic similarity among maize inbred lines detected by RFLPs, RAPDs, SSRs, and AFLP, *Theor. Appl. Genet.* **97**, 1248-1255.
- Peng, Y., Schertz, K.F., Cartinhour, S., and Hart, G.E. (1999) Comparative genome mapping of *Sorghum bicolor* (L.) Moench using an RFLP map constructed in a population of recombinant inbred lines, *Plant Breed.* **118**, 225-235.
- Pereira, M.G., Lee, M., Bramel-Cox, P., Woodman, W., Doebley, J., and Whitkus, R. (1994) Construction of an RFLP map in sorghum and comparative mapping in maize, *Genome* **37**, 236-243.
- Pereira, M., and Lee, M. (1995) Identification of genomic regions affecting plant height in sorghum and maize, *Theor. Appl. Genet.* **90**, 380-388.
- Pereira, M.G., Ahnert, D., Lee, M., and Klier, K. (1995) Genetic mapping of Quantitative Trait Loci for panicle characteristics and seed weight in Sorghum, *Revista Brasileira de Genetica* **18**, 249-257.
- Pernet, A., Hoisington, D., Dittinger, J., Jewell, D., Jiang, C., Khairallah, M., Letourmy, P., Marchand, J.L., Glaszmann, J.C., and González de León, D. (1999) Genetic mapping of maize streak virus resistance from the Mascarene source. II. Resistance in line CIRAD390 and stability across germplasm, *Theor. Appl. Genet.* **99**, 540-553.
- Perretant, M.R., Cadalen, T., Charmet, G., Sourdille, P., Nicolas, P., and *et al.* (2000) QTL analysis of bread-making quality in wheat using a doubled haploid population, *Theor. Appl. Genet.* **100**, 1167-1175.
- Philipp, U., Wehling, P. and Wricke, G. (1994) A linkage map of rye, *Theor. Appl. Genet.* **88**, 243-248.

- Plaschke, J., Börner, A., Xie, D.X., Koebner, R.M.D., Schlegel, R., and Gale, M.D. (1993) RFLP-mapping of genes affecting plant height and growth habit in rye, *Theor. Appl. Genet.* **85**, 1049-1054.
- Plaschke, J., Korzun, V., Koebner, R.M.D., and Börner, A. (1995) Mapping of the GA₃-insensitive dwarfing gene *ctl* on chromosome 7R in rye, *Plant Breed.* **114**, 113-116.
- Portyanko, V.A., Hoffman, D.L., Lee, M. and Holland, J.B. (2000) Molecular map of cultivated oat based on grass anchor DNA clone set, *Proc. Plant & Animal Genome Conference*, S. Diego, January 9-12, 2000, p. 111.
- Powell, W., Thomas, W.T.B., Baird, E., Lawrence, P., Booth, A., Harrower, B., McNicol, J.W., and Waugh, R. (1997) Analysis of quantitative traits in barley by the use of amplified fragment length polymorphisms, *Heredity* **79**, 48-59.
- Pratchett, N., and Laurie, D.A. (1994) Genetic map location of the barley developmental mutant liguleless in relation to RFLP markers, *Hereditas* **120**, 35-39.
- Price, A.H., Steele, K.A., Moore, B.J., Barraclough, P.B., and Clark, L.J. (2000) A combined RFLP and AFLP linkage map of upland rice (*Oryza sativa* L.) used to identify QTLs for root-penetration ability, *Theor. Appl. Genet.* **100**, 49-56.
- Qi, X., and Lindhout, P. (1997) Development of AFLP markers in barley, *Mol. Gen. Genet.* **254**, 330-336.
- Qi, X., Jiang, G., Chen, W., Niks, R.E., Stam, P., and Lindhout, P. (1999) Isolate-specific QTLs for partial resistance to *Puccinia hordei* in barley, *Theor. Appl. Genet.* **99**, 877-884.
- Qi, X., Stam, P., and Lindhout, P. (1998) Use of locus-specific AFLP markers to construct a high-density molecular map in barley, *Theor. Appl. Genet.* **96**, 376-384.
- Ragab, R.A., Dronavalli, S., Saghai Maroof, M.A., and Yu, Y.G. (1994) Construction of a sorghum RFLP linkage map using sorghum and maize DNA probes, *Genome* **37**, 590-594.
- Rami, J.F., Dufour, P., Trouche, G., Fliedel, G., Mestres, C., Davrieux, F., Blanchard, P., and Hammon, P. (1998) Quantitative trait loci for grain quality, productivity, morphological and agronomical traits in sorghum (*Sorghum bicolor* L. Moench.), *Theor. Appl. Genet.* **97**, 605-616.
- Rayapaty, P.J., Gregory, J.W., Lee, M., and Wise, R.P. (1994) A linkage map of diploid *Avena* based on RFLP loci and a locus conferring resistance to nine isolates of *Puccinia coronata* var. *Avenae*, *Theor. Appl. Genet.* **89**, 831-837.
- Reiter, R.S., Coors, J.G., Sussman, M.R., and Gableman, W.H. (1991) Genetic analysis of tolerance to low-phosphorus stress in maize using restriction fragment polymorphisms, *Theor. Appl. Genet.* **82**, 561-568.
- Ribaut, J.M., Hoisington, D.A., Deutsch, J.A., Jiang, C., and González-de-León, D. (1996) Identification of quantitative trait loci under drought conditions in tropical maize. I. Flowering parameters and the anthesis-silking interval, *Theor. Appl. Genet.* **92**, 905-914.
- Röder, M.S., Korzun, V., Wendehake, K., Plaschke, J., Tixier, M.-H., Leroy, P., and Ganal, M.W. (1998) A microsatellite map of wheat, *Genetics* **149**, 2007-2023.
- Ronald, P.C., Albano, B., Tabien, R., Abenes, L., Wu, K., and *et al.* (1992) Genetic and physical analysis of the rice bacterial blight resistance locus, Xa-21, *Mol. Gen. Genet.* **236**, 113-120.
- Russell, J.R., Fuller, J.D., Macaulay, M., Hatz, B.G., Jahoor, A., Powell, W., and Waugh, R. (1997) Direct comparison of levels of genetic variation among barley accessions detected by RFLPs, AFLPs, SSRs, and RAPDs, *Theor. Appl. Genet.* **95**, 714-722.
- Saghai Maroof, M.A., Yue, Y.G., Xiang, Z.X., Stromber, E.L., and Rufener, G.K. (1996) Identification of quantitative trait loci controlling resistance to gray leaf spot disease in maize, *Theor. Appl. Genet.* **93**, 539-546.
- Saito, A., Yano, M., Kishimoto, N., Nakagahra, M., Yoshimura, A., and *et al.* (1991) Linkage map of restriction fragment length polymorphism loci in rice, *Jpn. J. Breed.* **41**, 665-670.
- Sanchez, A.C., Ilag, L.L., Yang, D., Bler, D.S., Ausubel, F., Klush, G.S., Yano, M., Sasaki, T., Li, Z., and Huang, N. (1999) Genetic and physical mapping of *xa13*, a recessive bacterial blight resistance gene in rice, *Theor. Appl. Genet.* **98**, 1022-1028.
- Sari-Gorla, M., Krajewski, P., Binelli, G., Frova, C., Taramino, G., and Villa, M. (1997) Genetic dissection of herbicide tolerance in maize by molecular markers, *Molecular Breed.* **3**, 481-493.
- Sasaki, T. (1998) The rice genome project in Japan, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 2027-2028.
- Schachermayr, G., Siedler, H., Gale, M.D., Winzeler, H., Winzeler, M., and Keller, B. (1994), Identification and localization of molecular markers linked to the *Lr9* leaf rust resistance gene of wheat, *Theor. Appl. Genet.* **88**, 110-115.
- Schön, C.C., Melchinger, A.E., Boppenmaier, J., Brunklaus-Jung, E., Herrmann, R.G., and Seitzer, J.F. (1994) RFLP mapping in maize: quantitative trait loci affecting testcross performance of elite European flint lines, *Crop Sci.* **34**, 378-389.
- Senft, P., and Wricke, G. (1996) An extended genetic map of rye (*Secale cereale* L.), *Plant Breed.* **115**, 508-510.
- Senior, M.L., Chin, E.C.L., Lee, M., Smith, J.S.C., and Stuber, C.W. (1997) Simple sequence repeat markers developed from maize sequences found in the GENBANK database: map construction, *Crop Sci.* **36**, 1767-1683.

- Sherman, J.D., Fenwick, A.L., Namuth, D.M., and Lapitan, N.L.V. (1995) A barley RFLP map: alignment of three barley maps and comparisons to gramineae species, *Theor. Appl. Genet.* **91**, 681-690.
- Shizuya, H., Birren, B., Kim, U.J., Mancino, V., Slepak, V., Tachiiri, Y., and Simon, M. (1992) Cloning and stable maintenance of 300-kilobase-pair fragments of human DNA in *Escherichia coli* using an F-factor-based vector, *Proc. Natl. Acad. Sci. USA* **89**, 8794-8797.
- Siripoonwiwat, W., O'Donoghue, L.S., Wesenberg, D., Hoffman, D.L., Barbosa-Neto, J.F., and Sorrells, M.E. (1996) Chromosomal regions associated with quantitative traits in oat, *Quant. Trait Loci* **2**, Article 3.
- Song, W.H., Wang, G.L., Chen, L.L., Kim, H.S., Pi, L.Y., Holsten, T., Gardner, J., Wang, B., Zhai, W.X., Zhu L.H., Fauquet, C., and Ronald, P. (1995) A receptor kinase like protein encoded by the rice disease resistance gene, *Xa21*, *Science* **270**, 1804-1806.
- Springer, P.S., Edwards, K.J., and Bennetzen, J.L. (1994) DNA class organization on maize Adh1 yeast artificial chromosomes, *Proc. Natl. Acad. Sci. USA* **91**, 863-867.
- Stuber, C.W., Edwards, M.D., and Wendel, J.F. (1987) Molecular-marker-facilitated investigations of quantitative trait loci in maize. II. Factors influencing yield and its component traits, *Crop Sci.* **27**, 639-648.
- Stuber, C.W., Goodman, M.M., Schaffer, H.E., and Wier, B.S. (1980) Allozyme frequency changes associated with selection for increased grain yield in maize (*Zea mays* L.), *Genetics* **95**, 225-236.
- Stuber, C.W., Lincoln, S.E., Wolff, D.W., Helentjaris, T., and Lander, E.S. (1992) Identification of genetic factors contributing to heterosis in a hybrid from two elite maize inbred lines using molecular markers, *Genetics* **132**, 823-839.
- Stuber, C.W. (1995) Mapping and manipulating quantitative traits in maize, *Trends in Genetics* **11**, 477-481.
- Subudhi, P.K., and Nguyen, H.T. (2000) Linkage group alignment of sorghum RFLP maps using a RIL mapping population, *Genome* **43**, 240-249.
- Sun, G., Fahima, T., Korol, A., Turpeinen, T., Grama, A., Ronin, Y., and Nevo, E. (1997) Identification of molecular markers linked to the *Yr15* stripe rust resistance gene of wheat originated in wild emmer wheat, *Triticum dicoccoides*, *Theor. Appl. Genet.* **95**, 622-628.
- Tang, D., Wu, W., Li, W., Lu, H., and Worland, A.J. (2000) Mapping of QTLs conferring resistance to bacterial leaf streak in rice, *Theor. Appl. Genet.* **101**, 286-291.
- Tanoue, H., Shimokawa, T., Wu, J., Sue, N., Umehara, Y., and *et al.* (1997) Ordered YAC clone contigs assigned to rice chromosome 3 and 11, *DNA Res.* **4**, 133-140.
- Taramino, G., and Tingey, S. (1996) Simple sequence repeats for germplasm analysis and mapping in maize, *Genome* **39**, 277-287.
- Tautz, D. (1989) Hypervariability of simple sequences as a general source for polymorphic DNA markers, *Nucleic Acids Res.* **17**, 6463-6471.
- Thomas, W.T.B., Powell, W., Waugh, R., Chalmers, K.J., Barua, U.M., Jack, P., Lea, V., Forster, B.P., Swanston, J.S., Ellis, R.P., Hanson, P.R., and Lance, R.C.M. (1995) Detection of quantitative trait loci for agronomic yield, grain and disease characters in spring barley (*Hordeum vulgare* L.), *Theor. Appl. Genet.* **91**, 1037-1047.
- Thomas, W.T.B., Powell, W., Swanston, J.S., Ellis, R.P., Chalmers, K.J., Barua, U.M., Jack, P., Lea, V., Forster, B.P., Waugh, R., and Smith, D.B. (1996) Quantitative trait loci for germination and malting-quality characters in a spring barley cross, *Crop Sci.* **36**, 265-273.
- Tripathy, J.N., Zhang, J., Robin, S., Nguyen, Th.T., and Nguyen, H.T. (2000) QTLs for cell-membrane stability mapped in rice (*Oryza sativa* L.) under drought stress, *Theor. Appl. Genet.* **100**, 1197-1202.
- Tuberosa, R., Sanguineti, M.C., Landi, P., Salvi, S., Casarini, E., and Conti, S. (1998) RFLP mapping of quantitative trait loci controlling abscisic acid concentration in leaves of drought-stressed maize (*Zea mays* L.), *Theor. Appl. Genet.* **97**, 744-755.
- Tuinstra, M.R., Grote, E.M., Goldsbrough, P.B., and Ejeta, G. (1996) Identification of quantitative trait loci associated with pre-flowering drought tolerance in sorghum, *Crop Sci.* **36**, 1337-1344.
- Umehara, Y., Miyazaki, A., Tanoue, H., Yasukochi, Y., Saji, S., Otsuki, Y., and *et al.* (1995) Construction and characterization of rice YAC library for physical mapping, *Plant Breed.* **1**, 79-89.
- Van Deynze, A.E., Dubcovsky, J., Gill, K.S., Sorrells, M.E., Dvorak, J., Gill, B.S., Lagudah, E.S., McCouch S.R., and Appels, R. (1995a) Molecular-genetic maps for chromosomes 1 in *Triticeae* species and their relation to chromosomes in rice and oats, *Genome* **38**, 47-59.
- Van Deynze, A.E., Nelson, J.C., Yglesias, E.S., Harrington, S.E., Braga, D.P., McCouch, S.R., and Sorrells, M.E. (1995b) Comparative mapping in grasses. Wheat relationships, *Mol. Gen. Genet.* **248**, 744-754.
- Van Deynze, A.E., Nelson, J.C., O'Donoghue, L.S., Ahn, S.N., Siripoonwiwat, W., and *et al.* (1995c) Comparative mapping in grasses. Oat relationships, *Mol. Gen. Genet.* **249**, 349-356.
- Veldboom, L.R., Lee, M., and Woodman, W.L. (1994) Molecular marker-facilitated studies in an elite maize population: I. Linkage analysis and determination of QTL for morphological traits, *Theor. Appl. Genet.* **88**, 7-16.
- Vos, P., Hogers, R., Bleeker, M., Reijmans, M., van de Lee, T., Hornes, M., Fritters, A., Pot, J., Peleman, J., Kuiper, M., and Zabeau, M. (1995) AFLP: a new concept for DNA fingerprinting, *Nucleic Acids Res.* **23**, 4407-4414.

- Voylokov, A.V., Korzun, V., and Börner, A. (1998) Mapping of three self-fertility mutations in rye (*Secale cereale* L.) by using RFLP, isozyme and morphological markers, *Theor. Appl. Genet.* **97**, 147-153.
- Vuylsteke, M., Mank, R., Antonise, R., Bastiaans, E., Senior, M.L., Stuber, C.W., Melchinger, A.E., Lübberstedt, T., Xia, X.C., Stam, P., Zabeau, M., and Kuiper, M. (1999) Two high-density AFLP® linkage maps of *Zea mays* L.: analysis of distribution of AFLP markers, *Theor. Appl. Genet.* **99**, 921-935.
- Wang, G., Mackill, D.J., Bonman, J.M., McCouch, S.R., Champoux, M.C., and *et al.* (1994) RFLP mapping of genes conferring complete and partial resistance to blast in a durably resistant rice cultivar, *Genetics* **136**, 1421-1434.
- Wang, Z.M., Devos, K.M., Liu, C.J., Wang, R.Q., and Gale, M.D. (1998) Construction of RFLP-based maps of foxtail millet, *Setaria italica* (L.) P. Beauv., *Theor. Appl. Genet.* **96**, 31-36.
- Wang, Z.M., Yano, M., Yamamouchi, U., Iwamoto, M., Monna, L., Hayasaka, H., Katayose, Y., and Sasaki, T. (1999) The *Pib* gene for rice blast resistance belongs to the nucleotide binding and leucine-rich repeat class of plant disease resistance genes, *Plant J.* **19**, 55-64.
- Waugh, R., Bonar, N., Baird, E., Thomas, B., Graner, A., Hayes, P., and Powell, W. (1997) Homology of AFLP products in three mapping populations of barley, *Mol. Gen. Genet.* **255**, 311-321.
- Weber, D., and Helentjaris, T. (1989) Mapping RFLP loci in maize using B-A translocations, *Genetics* **121**, 583-590.
- Whitkus, R., Doebley, J., and Lee, M. (1992) Comparative genome mapping of sorghum and maize, *Genetics* **132**, 1119-1130.
- Williams, J.G.K., Kubelik, A.R., Livak, K.J., Rafalsky, J.A., and Tingey, S.V. (1990) DNA polymorphisms amplified by arbitrary primers are useful as genetic markers, *Nucleic Acids Res.* **18**, 6531-6535.
- Wu, J., Kurata, P., Tanoue, H., Shimokawa, T., Umehara, Y., Yano, M. and Sasaki, T. (1998) Physical mapping of duplicated genomic regions of two chromosome ends in rice, *Genetics* **150**, 1595-1603.
- Wu, P., Liao, C.Y., Hu, B., Yi, K.K., Jin, W.Z., Ni, J.J., and He, C. (2000) QTLs and epistasis for aluminium tolerance in rice (*Oryza sativa* L.) at different seedling stages, *Theor. Appl. Genet.* **100**, 1295-1303.
- Xie, D.X., Devos, K.M., Moore, G., and Gale, M.D. (1993) RFLP-based maps of the homoeologous group 5 chromosomes of bread wheat (*Triticum aestivum* L.), *Theor. Appl. Genet.* **87**, 70-74.
- Xu, G.W., Magill, C.W., Schertz, K.F., and Hart, G.E. (1994) RFLP linkage map of *Sorghum bicolor* (L.) Moench., *Theor. Appl. Genet.* **89**, 139-145.
- Xu, G.W., Subudhi, P.K., Crasta, O.R., Rosenow, D.T., Mullet J.E., and Nguyen H.T. (2000) Molecular mapping of QTLs conferring stay-green in grain sorghum, *Genome* **43**, 461-469.
- Yin, X., Kropff, M.J., and Stam, P. (1999) The role of ecophysiological models in QTL analysis: the example of specific leaf area in barley, *Heredity* **81**, 415-421.
- Yin, X., Stam, P., Dourleijn, J., and Kropff, M.J. (1999) AFLP mapping of quantitative trait loci for yield-determining physiological characters in spring barley, *Theor. Appl. Genet.* **99**, 244-253.
- Yoshimura, S., Yamanouchi, U., Katayose, Y., Toki, S., Wang, Z.X., Kono, I., Kurata, N., Yano, M., Iwata, N., and Sasaki, T. (1998) Expression of *Xa1*, a bacterial blight resistance gene in rice, is induced by bacterial inoculation, *Proc. Natl. Acad. Sci. USA* **95**, 1663-1668.
- Yu, Z., Mackill, D.J., Bonman, J.M., Tanksley, S.D. (1991) Tagging genes for blast resistance in rice via linkage to RFLP markers, *Theor. Appl. Genet.* **81**, 471-476.
- Yu, X.G., Bush, A.L., and Wise, R.P. (1996) Comparative mapping of homoeologous group 1 regions and genes for resistance to obligate biotrophes in *Avena*, *Hordeum* and *Zea mays*, *Genome* **39**, 155-164.
- Zanetti, S., Keller, M., Winzeler, M., Saurer, W., Keller, B., and Messmer, M. (1998) QTL for quality parameters for bread-making in a segregating wheat by spelt population, in A.E. Slinkard (ed.), *Proc. 9th Int. Wheat Genet. Symp.*, Extension Press, Saskatoon, Sask., Canada, **1**, 273-276.
- Zhang, H.B., and Wing, R.A. (1997) Physical mapping of the rice genome with BACs, *Plant Mol. Biol.* **35**, 115-127.
- Zheng, K., Shen, P., Quian, H., Wang, J. (1992) Tagging genes for wide compatibility in rice via linkage to RFLP markers, *Chinese Rice Sci.* **6**, 145-150.

GENE EXPRESSION UNDER ENVIRONMENTAL STRESSES - MOLECULAR MARKER ANALYSIS

Mikael Brosché¹, John R. Gittins¹, Helena Sävenstrand¹ and Åke Strid^{2,*}

¹*Biokemi och biofysik, Institutionen för kemi, Göteborgs universitet, P.O. Box 462, S-405 30 Göteborg, Sweden;* ²*Institutionen för naturvetenskap, Örebro universitet, S-70182 Örebro, Sweden;* **To whom correspondence should be addressed*

1. INTRODUCTION

Suboptimal environmental conditions affect plant performance both qualitatively and quantitatively. Qualitative changes are normally the result of protective responses by the plant in order to acclimate to the new circumstances and to withstand the threat posed by biotic or abiotic factors. Examples are accumulation of stress proteins (such as enzymes involved in detoxification of reactive oxygen species, components of biosynthetic pathways for secondary metabolites or polypeptides involved in the defence against invading organisms) and formation of substantial quantities of organic compounds which have different roles in detoxification and resistance (Dixon and Paiva, 1995; Kuc, 1995; Nocter and Foyer, 1998; Datta and Muthukrishnan, 1999). While conferring increased tolerance to the plant, the accumulation of such defensive compounds can possibly be toxic or distasteful for human beings or grazing livestock and so lead to the deterioration of crop quality (Caldwell *et al.*, 1994; Caldwell, 1998).

Quantitative changes, such as decreased growth rates, reduced crop yields or declining reproduction, are either the result of diversion of metabolic energy and photosynthate from the processes of cell division and elongation towards protection, or the result of direct effects on the components of catabolic and anabolic pathways. The action of environmental factors can thus have a severe negative economic impact on agriculture (Singh *et al.*, 1993; Hare *et al.*, 1996; Robertson *et al.*, 1996), horticulture and silviculture and it is consequently of great importance to achieve early detection for the implementation of measures to alleviate stress caused by biotic and abiotic agents.

To avoid large scale yield losses, it is expedient to use or to develop diagnostic molecular tools to monitor early changes in the metabolism in plants caused by environmental stresses, before morphological changes or decreases in growth patterns have occurred. The fact that many contemporary crop species are sown as monocultures with low genetic diversity and that varieties have been bred with the yield quantity and quality as the major criteria, has occasionally resulted in the loss of properties such as disease resistance, stress tolerance or adaptive capabilities in a changing environment (Reid and Miller, 1989). In the future, it can be envisaged that molecular stress detection will become a powerful tool for guiding in the application of agents possessing prophylactic or stress-alleviating properties.

There are several aspects to be considered when choosing a particular molecular marker for monitoring stress. Among these are the plant species to be studied, the stress specificity and the spatial and temporal pattern of expression of the chosen reporter gene, the cost, the facilities and equipment available, and training of staff to perform the test etc. The present review deals with techniques used to identify molecular markers for monitoring the effects on plants caused by anthropogenic environmental factors and we hope that it will serve as a guide for marker isolation depending on the stress problem to be tackled by the reader. This review will focus on markers for the detection of anthropogenic abiotic stresses including heavy metal exposure, exposure to gaseous compounds such as ozone, and UV-B radiation. Detection of non-anthropogenic abiotic stress (Winicov *et al.*, 2000) are dealt with in other chapters of this book. In addition, we will use the original definition of gene expression in describing the question at hand: "the process by which the information carried by a gene or genes becomes manifest as the phenotype" (Oxford dictionary of biochemistry and molecular biology, 1997), instead of only focusing on the detection at the nucleic acid level. However, activity measurements and the detection of accumulated metabolites will not be covered in this paper.

2. METHODS FOR FINDING DIFFERENCES IN GENE EXPRESSION UNDER ENVIRONMENTAL STRESS

2.1. Isolation of genes differentially regulated by environmental stress

2.1.1. Differential cDNA library screening

Methods to identify differentially expressed genes do so by comparing steady state levels of mRNA transcripts. Among the earliest methods used are the related procedures of differential cDNA library screening (this section) and library screening with subtracted probes (see section 2.1.3.). Both protocols require high quality cDNA libraries constructed using mRNA extracted from a plant tissue which has been subjected to the stress under investigation and from a tissue to be used as the control in the comparison of gene expression (usually the same type of tissue which has not been subjected to the stress). In addition, further stocks of mRNA from the stressed plant tissue and control tissue are needed for confirmation experiments (Table 1).

In differential cDNA library screening, the two mRNA pools are used separately as templates in cDNA synthesis reactions with one of the deoxynucleotide triphosphates replaced with its ^{32}P -labelled derivative. The mRNA templates are removed by treatment with alkali and then the two mixed cDNA probes are used in separate hybridisations to screen replicas of the stressed tissue cDNA library (Sambrook *et al.*, 1989). By examining the hybridisation pattern for differences, clones that are induced following stress can be selected for further study. If the differential screening is performed using a non-stressed control cDNA library this allows identification of genes that are repressed following stress treatment.

Differential screening is particularly suited to examining changes in gene expression in cells that have been subjected to stress or other stimuli since it detects quantitative differences in mRNA abundance. Although in principle this method can be used to identify any differentially regulated gene, it has been found to only work efficiently for genes that are highly expressed. The cDNA probes generated from total mRNA are highly complex because each cell can contain 20,000-30,000 or more different mRNAs (Bishop *et al.*, 1974; Axel *et al.*, 1976). Each cDNA species is represented in the probe mixture according to the abundance of the corresponding mRNA. Thus, the radioactivity incorporated in probe cDNAs representing rare transcripts will be too small to detect the cloned cDNA by library screening. It has been estimated that for the identification of a differentially expressed gene by

Table 1. Advantages and disadvantages of methods for identifying differentially expressed genes.

METHOD	ADVANTAGES	DISADVANTAGES
Differential library screening	<ul style="list-style-type: none"> Technically straightforward and does not require specialised equipment. Ideal for the identification of high level differentially expressed genes. 	<ul style="list-style-type: none"> A time consuming and indirect method. Individual clones identified by hybridisation have to be isolated before analysis. Only detects mRNAs whose abundance is significantly altered by a treatment. Only works well for highly expressed transcripts (>0.05% of total mRNA), unable to detect rare differentially expressed mRNAs.
Subtractive cDNA libraries	<ul style="list-style-type: none"> In principle, the method of choice to isolate as complete a set of differentially expressed genes as possible. cDNAs corresponding to rare mRNAs (<0.001%) are enriched. Reduces the number of clones to be screened with a subtracted probe by up to a factor of 10. 	<ul style="list-style-type: none"> Requires large amounts of starting mRNA. Technically difficult to construct a subtractive cDNA library. Low amounts of cDNA recovered after subtraction which may be degraded. Hydroxyapatite columns used to separate single stranded cDNA from hybrids are cumbersome and inefficient (the use of biotin/streptavidin and magnetic separation technology have overcome these problems) Screening the library with a subtracted probe is inefficient (see above). Time consuming and indirect. Can only compare the mRNA populations in two tissues at a time.
Library screening with a subtracted probe	<ul style="list-style-type: none"> Useful when there are few differences between the test and control mRNA populations. Suitable for the isolation of rare transcripts (0.005-0.01% of total) because sequences common to test and control mRNA populations are removed from the probe. The cDNA library to be screened can be plated at a higher density because mRNAs hybridising to abundant sequences are removed from the probe. 	<ul style="list-style-type: none"> Requires large amounts of starting mRNA to drive subtractive hybridisation to completion. Hydroxyapatite columns used to separate single stranded cDNA from hybrids are cumbersome and inefficient (the use of biotin/streptavidin and magnetic separation systems have overcome this problem). Subtraction rapidly reduces the amounts of cDNA in the probe and makes it difficult to perform multiple rounds. Low levels of radioactivity in the subtracted probe for each given cDNA requires prolonged exposure of autoradiographs. Time consuming and indirect. Unable to detect mRNAs which are present in both cell types but more abundant in one than the other.

Table 1 continued

METHOD	ADVANTAGES	DISADVANTAGES
Differential display RT-PCR	<ul style="list-style-type: none"> • In principle, can permit the identification of all expressed genes. • Satisfactory method if only a small number of differentially expressed molecular markers are required. • The use of PCR reduces the requirement for the amount of starting mRNA. • A relatively rapid procedure. • Many tissues can be compared simultaneously. • Both up- and down-regulation of genes may be detected at the same time. 	<ul style="list-style-type: none"> • Requires the use of expensive reagents (radioactive isotope/Taq polymerase) and equipment (thermal cycler for PCR). • The handling of a lots of radioactivity (^{32}P) and neurotoxic acrylamide is necessary. • Not efficient for the isolation of rare mRNAs. Genes with moderate to high levels of expression are preferentially identified. • False positives can be a real problem because each band on the gel can contain multiple DNA fragments. • An indirect method.
cDNA-AFLP	<ul style="list-style-type: none"> • Materials and equipment used for differential display and/or genomic AFLP are directly applicable to cDNA AFLP. • Due to the highly stringent PCR conditions the method is reproducible and false positives are less likely. • The method is not as biased towards abundant transcripts compared to differential display and so rare mRNA species may be identified. • Only perfect primer-template matches are amplified and this should simplify the process of cloning differentially expressed bands from the gel "fingerprint". • In one variation of the method, regions of the cDNA that come within the translated portion are isolated rather than sequences from the 3' end, which facilitates gene identification. • When coupled with nuclear-expressed sequence tags the method can give precise information about gene expression profiles in specific cell populations. 	<ul style="list-style-type: none"> • Only cDNAs that are cleaved by both the restriction endonucleases used in the protocol are included in the analysis. This may mean that as much as half of all transcripts are omitted. • Similarly to differential display, co-migration of PCR products on the sequencing gel might contribute to false positives. • The method has received little attention to date and so its use has yet to be fully explored.

Table 1 continued

METHOD	ADVANTAGES	DISADVANTAGES
Suppression subtractive hybridisation PCR (SSH)	<ul style="list-style-type: none"> • An improvement on the generation of subtractive cDNA libraries, overcoming most of the problems. • The use of PCR reduces the requirement for the amount of starting mRNA. • PCR also permits multiple rounds of subtraction so that extremely rare transcripts may be identified (2-3 fold differential expression). • A rapid procedure. • Screening of the subtracted library using a subtracted probe is unnecessary. A high proportion of clones contain differentially expressed cDNAs (25-100%). 	<ul style="list-style-type: none"> • Requires the use of expensive reagents (magnetic separation technology/ <i>Taq</i> polymerase) and equipment (thermal cycler for PCR). • Although a direct method, the resulting differentially expressed cDNAs are only small fragments. Further effort is required to isolate full length clones. • Can only compare the mRNA populations in two tissues at a time. • Like all subtraction methods, is unable to identify differentially expressed genes encoding polypeptides with domains sharing homology to those of constitutively expressed proteins.
Total cDNA library sequencing (ESTs)	<ul style="list-style-type: none"> • A simple method, just sequence two cDNA libraries and compare them. • The ultimate way of identifying all differentially expressed genes. • Suitable for automation. • Can be coupled to DNA array technology to examine coordinated gene expression following stress. 	<ul style="list-style-type: none"> • Requires massive investment in materials, laboratory equipment, computer facilities and personnel. • Unless mRNA populations are quite different, the identification of differentially expressed genes is not more rapid than suppression subtractive hybridisation PCR. • Not efficient for the isolation of rare mRNAs unless normalised and/or subtracted libraries are sequenced.
Serial analysis of gene expression (SAGE)	<ul style="list-style-type: none"> • A means of both quantitative and qualitative analysis of an mRNA population. • Once SAGE "libraries" have been constructed, only DNA sequencing is required to obtain extensive expression profiles of high resolution. • A potential source of many differentially expressed markers. • Improvements to the original method will make the technique less difficult, applicable to situations where starting material is scarce and able to provide more information about genes from which SAGE tags originate. 	<ul style="list-style-type: none"> • Constructing of SAGE "libraries" is a technically difficult multi-step process with many opportunities for human error and losses in material. • Standard procedure requires large amounts of starting mRNA. • For a quantitative picture of gene expression, many SAGE tags need to be sequenced (eg. >10,000) with the associated cost and effort involved. • SAGE tag sequence data must be carefully screened to remove sequences over represented due to preferential amplification, adapter/linker contamination or other PCR artifacts. • Due to the small size of the SAGE tags and because they are often derived from the untranslated 3' end of mRNAs, many will have no match in EST sequence databases or they may match multiple sequences from unrelated genes. • SAGE preferentially identifies abundant mRNAs and is thus not suited to the rapid identification of rare transcripts. • An indirect method.

this method, the mRNA must represent approximately 0.05% of the total mRNA in one cell type and less than 0.01% in the other (Sambrook *et al.*, 1989).

The use of differential library screening for the identification of stress-specific genes in plants has been widespread. Changes in gene expression caused by raised levels of the anthropogenic abiotic stressors ozone (Eckey-Kaltenbach *et al.*, 1997, Wegener *et al.*, 1997), ultraviolet radiation (Toguri *et al.*, 1993), heavy metals (Didierjean *et al.*, 1996) and aluminium (Snowden and Gardner, 1993; Richards *et al.*, 1994; Cruz-Ortega *et al.*, 1995; Ezaki *et al.*, 1995; Richards and Gardner, 1995; Richards *et al.*, 1995; Ezaki *et al.*, 1996; Ezaki *et al.*, 1997; Richards *et al.*, 1998) have been examined by the use of this method (Table 2).

2.1.2. Subtractive cDNA libraries

Another method to find differentially expressed genes is subtractive cloning, which has been widely used in medical research (Sagerström *et al.*, 1997). This method has been used in several studies of gene expression at different stages of plant development such as flowering etc. (Buchanan-Wollaston and Ainsworth, 1997; Furuyama and Dzelzkalns, 1999). It has also been used to identify a number of different stress-induced genes (Ouvrard *et al.*, 1996).

Subtraction techniques involve the subtraction of mRNA from one population of cells from that of another. mRNA species that are present in both populations will be removed by hybridisation and only the differentially expressed mRNA will remain. The population of mRNAs from which differentially expressed genes are to be isolated is called the tester sample and the control sample is called the driver (Sagerström *et al.*, 1997).

Since mRNA molecules are unstable and easily degraded, cDNA is usually synthesised from the tester and the driver and used for the hybridisation reaction. The tester and the driver cDNAs are denatured and mixed and hybrids are formed. Tester-driver, driver-driver and tester-tester hybrids are created. To ensure that differentially expressed genes are selected, the driver should be present at a much higher concentration than the tester. A tester:driver ratio of 1:10 is commonly used. Another parameter controlling the hybridisation is whether the driver and the tester are single or double-stranded. Using single-stranded driver is the most efficient for hybridising to sequences shared by the two populations but a double-stranded driver may also be used although this may require additional rounds of subtraction. This is often done routinely even if using the single-stranded driver in order to reduce the number of false positives.

In the subtraction step, the cDNAs common to both populations of cDNA are subtracted by the removal of nucleic acid hybrids. Several methods have been developed for this purpose. Hydroxyapatite, which binds double-stranded nucleic acids, can be used to separate single-stranded unhybridised cDNA from the hybrids. The use of biotinylated driver, is another way to subtract the hybrid from the single-stranded tester. The driver-tester hybrids are then removed by a binding reaction to streptavidin. Other less frequently used methods are chemical crosslinking and degradation of driver-tester hybrids (Sagerström *et al.*, 1997). After a few rounds of hybridisation and subtraction, the remaining cDNAs are converted to double stranded form (where necessary) and ligated into a suitable vector and cloned for characterisation.

The original method has been further developed and refined over the years. By the application of PCR, it is now possible to amplify the differently expressed mRNAs to allow multiple rounds of subtraction against different driver samples (section 2.1.6.; Wang *et al.*, 1991; Balzer and Bäumlein, 1994). In spite of this, some disadvantages with the method remain (Table 1). Large amounts of mRNA are still needed to perform the subtraction, and due to the reaction kinetics, constitutively expressed but rare mRNA will often remain unhybridised (Sagerström *et al.*, 1997).

2.1.3. Screening cDNA libraries with subtracted probes

Instead of cloning the cDNAs after subtraction (i.e. constructing a subtractive cDNA library), the subtracted cDNAs labelled with ^{32}P can be used directly to screen an existing cDNA library. Hybridising clones are then isolated for further study.

Library screening with a subtracted probe is a method more suited for situations where there are very few differences between two mRNA populations (Table 1). It is most applicable when a small proportion of genes are not expressed at all in one tissue compared to another. This procedure then detects qualitative differences in mRNA abundance.

As an example of the use of this method, screening of a cDNA library constructed using mRNA from aluminium-treated soybean roots, using a subtracted probe, has been performed. Two related aluminium-induced genes were identified in this way (Ragland and Soliman, 1997).

2.1.4. Differential display and randomly arbitrarily primed PCR

The closely related techniques of differential display and randomly arbitrarily primed PCR (RAP-PCR) were both introduced in 1992 (Liang

and Pardee, 1992; Welsh *et al.*, 1992). The procedure begins with a cDNA synthesis from isolated total RNA or mRNA using an oligo-dT₁₂VN primer (V=A, C or G; N=A, C, G or T; differential display) or a random primer (RAP-PCR). This cDNA synthesis is performed with RNA isolated from all samples that are to be compared, e.g. stress-exposed and control tissues are used. A PCR reaction is then carried out by using the resulting cDNA as the template and the cDNA synthesis primer in combination with a second random primer. A radioactively labelled nucleotide is also included in the reaction. The resulting PCR products from the stressed and control samples are separated on a sequencing gel which is then dried and placed against X-ray film. On the autoradiograph, the banding patterns are compared to find bands present only during one of the conditions under investigation (e.g. stress-exposed or control samples). If such a band is found it is excised from the dried gel, reamplified by PCR and cloned into a vector. The PCR-product may then be used to probe a northern blot to verify the differential expression before it is sequenced.

Initially, differential display was thought to solve some of the problems present in earlier methods, such as poor sensitivity and the demand for large amounts of RNA. Another potential advantage was that with differential display both up- and down-regulated genes could be found in the same experiment. In practice, however, many laboratories have found that differential display is an inefficient method with severe problems mainly caused by a large number of false positives, i.e. bands thought to correspond to differentially regulated genes that turn out not to be so when used as probes on northern blots. This problem is time-consuming and creates a large demand for RNA for northern blots. A further problem is that the method is apparently biased towards abundant mRNAs at the expense of rare messages (Bertioli *et al.*, 1995). Despite these drawbacks, differential display has been successfully used to isolate differentially regulated genes following UV-irradiation and ozone exposure (Sharma and Davis, 1995; Kiiskinen *et al.*, 1997; Lers *et al.*, 1998; Brosché and Strid, 1999a; Brosché and Strid, 1999b; Brosché *et al.*, 1999; Table 2).

2.1.5. cDNA-AFLP and NEST

cDNA-AFLP or AFLP based mRNA fingerprinting is an extension of differential display that overcomes the selection against rare mRNA species and other drawbacks inherent in this method (Bachem *et al.* 1996; Money *et al.* 1996; Bachem *et al.* 1998; Ivashuta *et al.* 1999). The procedure differs from the standard differential display protocol in steps following the initial cDNA synthesis. Instead of amplification of 3'

transcript fragments using random primers of arbitrary sequence in combination with poly-dT primers in low annealing temperature PCR reactions, a stringent AFLP-PCR is used. Double-stranded cDNA is synthesised, digested with restriction endonucleases and “anchor” adaptors are ligated to the termini. “Anchor”-specific primers are then used in a pre-amplification reaction to generate larger quantities of template for subsequent reactions. This secondary template is used with a panel of AFLP primers (which are identical to the “anchor”-specific primers but contain specific base extensions at the 3' end) in PCR reactions with high annealing temperatures. To increase the selectivity of the amplification, one modification of the procedure uses suppression PCR (Fukuda *et al.* 1999). As with conventional differential display, the product bands can be analysed by denaturing polyacrylamide gel electrophoresis with autoradiography if one of the AFLP primers used in this stringent PCR is radiolabelled. Alternatively, sensitive silver staining may be used to detect the transcript-derived fragments (Ivashuta *et al.* 1999). By the use of different combinations of AFLP primers in separate reactions, many different specific bands can be amplified and displayed as an expression fingerprint. Expression profiles in different tissues or following different treatments can thus be compared as with differential display. The highly stringent PCR conditions used in cDNA-AFLP mean that only perfect primer-template matches are amplified and template quantity is less important; so the method is reproducible, false positives are less likely and rare differentially expressed mRNA species may be identified. Specific transcript derived fragments can be excised from the gels and reamplified using AFLP primers for further characterisation.

For cDNA-AFLP to work effectively, the choice of restriction endonucleases used to cleave the double-stranded cDNA is critical because this step determines the proportion of transcripts included in the analysis. In the method described by Bachem *et al.* (1996, 1998) a pair of restriction endonucleases are used, one with a 4 bp recognition sequence and the other with a 6 bp recognition sequence. Ideally both should cleave each cDNA at least once, however in practise approximately half of cDNAs are not cleaved by one of the enzymes (Bachem *et al.* 1998). Using this protocol, the amplified fragments are often derived from internal coding regions of cDNAs rather than the 3' ends and so the frequency with which specific transcripts can be identified is increased compared with conventional differential display. In the variation of cDNA-AFLP described by Money *et al.* (1996), an *Eco* RI restriction site is incorporated in the initial poly-dT cDNA synthesis primers and *Eco* RI is one of the enzymes used to cleave the double-stranded cDNA. This facilitates the use of standard genomic AFLP “anchor” adaptors and primers but has the drawback that the final amplified fragments originate from the 3' non-coding end of the cDNAs.

A recent extension of cDNA-AFLP is to combine it with a technique known as nuclear-expressed sequence tag (NEST) analysis (Macas *et al.* 1998). This method which involves analysis of nuclear transcripts has so far been used to compare transcription in different tissues and at different developmental stages, although its use to examine differential expression following abiotic stress can be envisaged. Transgenic plants expressing nuclear-targeted green fluorescent protein (GFP) under the control of a stress-responsive promoter and subjected to the relevant stress treatment would provide the starting material. Subcellular homogenates derived from these plants would be subjected to flow cytometry to sort and isolate nuclei containing GFP. By performing cDNA-AFLP using polyadenylated RNA purified from these sorted nuclei and comparing the profile with that produced using mRNA from nuclei of non-stressed plants, genes which are co-ordinately regulated following stress may be identified. This analysis would be more precise than a similar study performed without the nucleus sorting step because it would only consider stressed cells within a tissue.

Despite the apparent advantages of cDNA-AFLP and the potential of NEST, neither of these techniques has been widely used to date.

2.1.6. Suppression subtractive hybridisation

Suppression subtractive hybridisation (SSH) was developed from the subtraction cloning described above and has also been named PCR-select cDNA subtraction. The general principle is the same as for constructing subtractive cDNA libraries, i.e. subtraction of the mRNA of one population from another. However, in SSH, due to the use of PCR, the differentially expressed mRNA will be exponentially amplified (Table 1) and the synthesis of undesirable cDNA will be suppressed (Diatchenko *et al.*, 1996; Diatchenko *et al.*, 1999).

The method is performed as follows: 1) cDNA is synthesised from both populations of mRNA, the control mRNA (driver) and the test mRNA (tester). The two cDNA populations are fragmented by digestion with a 4-base recognition site restriction enzyme which produces blunt ends; 2) The tester cDNA is split into two aliquots and to each of these a unique adaptor is ligated; 3) The hybridisation is performed between half the amount of driver and one of the two tester populations. Driver-tester hybrids will be formed and only differentially expressed cDNAs will remain single-stranded; 4) This hybridisation is repeated with the second tester population and the remainder of the driver population. Again hybrids will be formed and only differentially expressed cDNA will remain single-stranded; 5) A second hybridisation is

Table 2. Differentially expressed genes identified using the various methods described.

A: Differential Library Screening				
Stress	Differentially expressed gene product	Species	Regulation	Reference
Hg ²⁺ UV supplemented white light	Glycine rich proteins, Heat shock protein, Thaumatin-like protein Pathogenesis related proteins, Ubiquitin Membrane proteins	Maize	Up-regulated	Didierjean <i>et al.</i> , 1996
	Cytochrome P-450	Eggplant	Up-regulated	Toguri <i>et al.</i> , 1993
Aluminium	Phenylalanine ammonia lyase Cysteine-rich metallothionein homologue Bowman-Birk proteinase inhibitor homologues	Wheat	Up-regulated	Snowden and Gardner, 1993
	Bowman-Birk proteinase inhibitor homologues, Unidentified ORF	Wheat	Up-regulated	Richards <i>et al.</i> , 1994
	I,3-beta-glucanase	Wheat	Up-regulated	Cruz-Ortega <i>et al.</i> , 1995
	Auxin regulated gene Glutathione S-transferase	Tobacco	Up-regulated	Ezaki <i>et al.</i> , 1995
	Proline-rich hydrophobic protein	<i>Arabidopsis</i>	Up-regulated	Richards and Gardner, 1995
	Proline-rich hydrophilic repeat protein	<i>Arabidopsis</i>	Up-regulated	Richards <i>et al.</i> , 1995
	Moderately anionic peroxidase	Tobacco	Up-regulated	Ezaki <i>et al.</i> , 1996
	GDP dissociation inhibitor	Tobacco	Up-regulated	Ezaki <i>et al.</i> , 1997
	Peroxidase, Glutathione S-transferase, Blue copper binding protein, CAB Reticuline:oxygen oxidoreductase homologue, Aldolase, Ala aminotransferase	<i>Arabidopsis</i>	Up-regulated	Richards <i>et al.</i> , 1998

Table 2 continued

B: Screening DNA libraries with subtracted probes					
Stress	Differentially expressed gene product	Species	Regulation	Reference	
Aluminium	Auxin down-regulated gene	Soybean	Up-regulated	Ragland and Soliman, 1997	
C: Differential Display					
Stress	Differentially expressed gene product	Species	Regulation	Reference	
Ozone	Mitochondrial phosphate translocator protein	Birch	Up-regulated	Kiiskinen <i>et al.</i> , 1997	
Ozone/Pathogen	Unidentified ORF				
Ozone/UV-B	Ribosomal protein S26	<i>Arabidopsis</i>	Up-regulated	Sharma and Davis, 1995	
Ultraviolet radiation (UV-C)	Isoflavone reductase-like protein	Pea	Up-regulated	Brosché and Strid, 1999a	
UV-B	Chlorophyll a/b-binding protein, Rubisco small subunit, PSII 33kDa protein, PSII D1 polypeptide	Grapefruit	Up-regulated	Lers <i>et al.</i> , 1998	
UV-B/ozone	SadA and SadC	Pea	Down-regulated	Brosché <i>et al.</i> , 1999	
		Pea	Up-regulated	Brosché and Strid, 1999b	
D: Suppression subtractive hybridisation (SSH) and other PCR based subtraction methods					
Stress	Differentially expressed gene product	Species	Regulation	Reference	
Cd ²⁺	Unidentified ORF	<i>Arabidopsis</i>	Up-regulated	Choi <i>et al.</i> , 1995	
Ozone	Pathogenesis related protein 4A, Extensin, Pre-hevein-like protein, Lipid transfer protein, Leucine rich repeat protein, Disease resistance response protein, Unidentified ORF	Pea	Up-regulated	Sävenstrand <i>et al.</i> , 2000	

performed and the two primary hybridisation mixtures pooled without any further denaturation. Now only the single-stranded cDNAs from the two initial hybridisations will hybridise; 6) The special hybrids formed by differentially expressed cDNAs have different adaptors ligated to each end and due to this fact the cDNAs can easily be amplified with PCR; 7) The PCR-products can then be directly inserted into a cloning vector and analysed.

The advantage with SSH compared to other methods are that a smaller amount of mRNA is needed for the reaction and that the PCR-amplification will increase the chances of finding rare mRNAs (Diatchenko *et al.*, 1996; Table 1).

Despite all the advantages of SSH, the method has not been widely used when searching for molecular markers in plants. Only two reports have appeared in which SSH or similar methods have been used to identify genes induced by abiotic stress; one describes an unidentified gene induced by cadmium stress in *Arabidopsis* (Choi *et al.*, 1995) and another describes several genes up-regulated in pea plants following ozone exposure (Sävenstrand *et al.*, 2000; Table 2). SSH has also been used to identify genes induced during flower development (Kim *et al.*, 1999).

2.1.7. Total cDNA library sequencing: ESTs

For a definitive description of transcript accumulation following stress, a comparison of saturated EST (expressed sequence tag) collections from the target tissue, before and after stress, is required (Table 1). In practice, this means constructing separate high quality cDNA libraries from the stressed and control tissues and performing single pass DNA sequencing to obtain sequence data from either end of randomly selected clones. Each new EST is compared to sequences in the existing databases to assign putative functions to each clone. From the large collection of sequence data obtained, genes which are induced or repressed by the stress under study can be identified by quantifying their abundance in the sequenced EST populations, a process which has been called electronic subtraction (Wan *et al.*, 1996). This designation can be confirmed individually by using the selected clones as northern blot probes or more efficiently by constructing EST microarrays (see section 3.1.4) which may be hybridised with probes derived from test and control mRNAs.

To obtain a saturated EST collection, it may be necessary to sequence over 100,000 clones. This task would have been impractical 10 years ago because of the manual labour involved, as well as the cost, but with the recent implementation of new sequencing technologies, automated sample processing and new computer facilities for data handling and analysis, it is now feasible. However even with this large

number of sequences it is still possible that rare mRNAs will be missed. If the pattern of mRNA complexity in plants is similar to that seen in mammalian systems, the majority of transcripts within the cell represent between $1/20,000$ - $1/70,000$ of the total population, and the rarest mRNAs are even less abundant (Bishop *et al.*, 1974; Axel *et al.*, 1976). It is therefore prudent to sequence additional clones from normalised and/or subtracted libraries to gain a more complete picture of the transcript population and to identify rare clones. If the goal is simply to identify a few marker genes which can be used as expression reporters of stress, particularly where the test and control mRNA populations are very similar, total cDNA library sequencing will not produce more rapid results than a method such as SSH.

There are a growing number of plant EST projects throughout the world examining a broad spectrum of species. Besides the well established major efforts on *Arabidopsis* (Cooke *et al.*, 1996; Rounsley *et al.*, 1996) and rice (Sasaki *et al.*, 1994; Yamamoto and Sasaki, 1997), there are large ongoing collaborative projects on commercially important species including *Brassica napus*, loblolly pine, maize, potato, soybean, tomato and wheat. There are also many other smaller projects on diverse species, the results of which are catalogued in the database of ESTs (<http://www.ncbi.nlm.nih.gov/dbEST>). Gene expression in plants following different stresses is a component of some of these projects and results can be expected in the near future.

2.1.8. Serial analysis of gene expression (SAGE)

Another recent development for the identification of differentially expressed genes is serial analysis of gene expression (SAGE). This technique, which was devised by Velculescu *et al.*, (1995), is a means of both quantitative and qualitative analysis of an mRNA population. To perform SAGE, short cDNA nucleotide tag sequences (usually 10 bp) are generated from immobilised cDNA. Sequences of this length usually contain sufficient information to identify the gene from which they were derived. These tags are linked together in such a way that the boundary between each can be distinguished and the concatamers are then cloned and sequenced. The frequency of a specific tag in these SAGE "libraries" is indicative of the expression level of the gene. Once SAGE data has been compiled the profiles of gene expression in two different tissues (e.g. before and after stress) can be compared to identify differentially expressed tags and the corresponding genes.

Despite the elegance of SAGE (Table 1), the current procedure has been applied mainly to yeast and human situations (see <http://www.sagenet.org/>), with few studies involving plants (Matsumura *et*

al., 1999). There are a number of reasons why SAGE has been slow to gain widespread popularity: (1) The primary reason is the difficulty experienced by some workers in constructing SAGE "libraries". The procedure is technically difficult requiring many sequential enzymatic reactions with purification and concentration steps in between. As well as presenting many opportunities for human error, the multi-step procedure can lead to significant losses in material; (2) The method requires a large amount of input mRNA (eg. 2.5-5.0 μg polyA⁺ RNA) which may be difficult to obtain from some tissues. However, a new modification called microSAGE, which requires only 1-5 ng polyA⁺ RNA (Datson *et al.*, 1999) may make the technique more widely applicable; (3) When SAGE is to be used to get a quantitative picture of gene expression, many tags need to be sequenced (>10,000). This requires a significant amount of DNA sequencing with the associated cost and effort involved; (4) All SAGE tag sequence data must be carefully screened to remove sequences over-represented due to preferential amplification, and tags produced by adapter/linker contamination or other PCR artifacts; (5) Due to the small size of the SAGE sequence tags, and the fact that they are frequently derived from the untranslated 3' end of mRNAs, many of them will have no match in EST sequence databases (particularly when a less common plant species is being studied) or they may match multiple sequences from unrelated genes. Using the tag sequences with oligo(dT) as the primers and cDNA as the template, PCR may be used to obtain more sequence information from the gene(s) containing the tags (Chen *et al.*, 2000); (6) SAGE "libraries" are not normalised and so the method preferentially identifies abundant mRNAs and is thus not suited to the rapid identification of rare transcripts.

Despite all these potential drawbacks, once good SAGE "libraries" have been generated and exhaustively sequenced, it is possible to obtain extensive expression profiles of high resolution which can be the source of many differentially expressed markers. In the near future, the development of commercial kits which simplify the construction of SAGE "libraries" from small amounts of starting material combined with the use of improved computer programs to compile SAGE data and remove artifacts, will make this powerful technique more popular among plant researchers.

2.2. Expression screens for stress-related proteins, activities, protein-protein interactions and protein-DNA interactions

The true expression of a particular gene is actually determined by the amount of functional protein that is produced after induction of a gene and not, as sometimes is anticipated, the steady state concentration of

the particular mRNA in a cell, a tissue or an organism. A rapid and massive increase in an mRNA does not necessarily correspond to an equivalent accumulation of the corresponding active enzyme. The flow of information through the components of the central dogma, from the gene, through transcription via a primary transcript, to mature polypeptide, is affected by processes other than the rate of transcription. Among these are post-transcriptional processes such as mRNA degradation or splicing of primary transcripts (Gutiérrez *et al.*, 1999; Lorkovic *et al.*, 2000), selective translation of mRNAs (Bailey-Serres, 1999) and post-translational events such as covalent modification (Edelman *et al.*, 1987; Clark, 1993; Lodl, 1994; O'Farrell, 1995; Dupree and Sherrier, 1998; Pirrung, 1999; Rodriguez-Concepcion *et al.*, 1999), intracellular trafficking (Bar-Peled *et al.*, 1996), and proteolytic cleavage of the protein (Jung *et al.*, 1998).

Thus, to actually prove that a change in the plant's environment leads to a change in gene expression, measurements at the protein level are necessary, such as identification of novel unknown proteins, estimation of the cellular content of a desired marker, or measurement of the activity of the interesting protein, be it a catalytic reaction or an activity of binding to smaller molecules or to other large biomolecules.

Thus, to actually prove that a change in the plant's environment leads to a change in gene expression, measurements at the protein level are necessary, such as identification of novel unknown proteins, estimation of the cellular content of a desired marker, or measurement of the activity of the interesting protein, be it a catalytic reaction or an activity of binding to smaller molecules or to other large biomolecules.

2.2.1. Immunological techniques

In the case where a molecular marker protein is known, has been isolated and purified, and antibodies (polyclonal or monoclonal) have been raised against it, the actual detection in tissues exposed to environmental stress is straight-forward. The presence of the protein in tissue extracts can be analysed and quantified by using standard immunological techniques, such as western blotting and immunolabelling or ELISA (Coligan *et al.*, 1997). Most studies of molecular markers at the protein level aim at producing antibodies in order to capitalise on these simple and cost-effective methods. However, in cases where novel proteins are to be investigated or when pure preparations of polypeptide from the original source or from a recombinant over-expressing system cannot be used as antigens in antibody production, other means of detection and identification have to be used.

2.2.2. Finding novel proteins to be used as molecular markers

2-D electrophoresis and mass spectrometry analysis. Recent improvements in the resolution and reproducibility of 2-D electrophoresis, combined with developments in biomolecule ionisation techniques applicable to mass spectrometry, now permit the systematic detection and identification of proteins present in the cell (Fichmann and Westermeier, 1999). As an extension of this procedure, changes in the levels of proteins following different stress treatments may be quantified by incorporation of stable nuclids, such as ^{15}N , into the proteins of one of the cell populations to be compared (Mann, 1999).

In 2-D electrophoresis (Fichmann and Westermeier, 1999), the proteins in a sample are first separated according to their isoelectric point (by isoelectric focusing) in the first dimension followed by separation according to their size in the second dimension (with SDS-PAGE). The gels may then be stained with coomassie brilliant blue or silver, or blotted onto membranes for detection with antibodies. Several different computer programmes are available to analyse the resulting protein spot pattern (Pleißner *et al.*, 1999). A good starting point for a researcher interested in 2-D electrophoresis may be found on the ExPASy molecular biology server (<http://www.expasy.ch/>).

Once an interesting spot has been identified by 2-D electrophoresis, it can be excised from the gel and its identity determined by mass spectrometry (Courchesne and Patterson, 1999). Prior *in situ* treatment of the individual protein spots with sequence-specific proteases such as trypsin produces the best starting material for this type of analysis (Shevchenko *et al.*, 1996). Peptide mapping is then achieved by using matrix-assisted laser desorption ionisation (MALDI) coupled to a time-of-flight (TOF) mass spectrometer. The resulting mass fingerprint then can be used together with information in existing protein and nucleic acid databases (such as SwissProt, GenBank or different EST databases) for computational identification of each protein analysed (Eng *et al.*, 1994; Mann and Wilm, 1994; Zhang and Chait, 2000). When sequence data are only fragmentary or unavailable, MALDI-TOF analysis should be complemented with further analysis of interesting fragments by using an electrospray ionisation quadrupole time-of-flight spectrometer (ES Q-TOF MS) in order to reveal the amino acid sequences of the peptides (Andersen and Mann, 2000).

It is likely that novel developments in the area of mass spectrometry will soon make it possible to achieve simultaneous peptide mapping and sequence determination by MALDI hybrid quadrupole time-of-flight MS (Krutchinsky *et al.*, 2000). Furthermore, several different methods for mass spectrometry analysis of complex mixtures of proteins without prior 2-D electrophoresis are being developed. Some of these

exploit new or improved separation techniques (e.g. capillary reverse phase liquid chromatography or isoelectric focusing) coupled to diverse mass spectrometry methods such as Fourier transform ion cyclotron resonance mass spectrometry (Jensen *et al.*, 1999) or tandem mass spectrometry (Li *et al.*, 1999). Alternatively, protein digests may be analysed directly by liquid chromatography mass spectrometry (Link *et al.*, 1999).

For recent reviews on mass spectrometry as a tool for proteomics, including short descriptions of the different approaches, see Pandey and Mann (2000) or Andersen and Mann (2000). Information on computer analysis tools for protein identification from mass spectrometry data can for instance be found at <http://prospector.ucsf.edu/ucsfhtml3.2/mstagfd.htm> or at <http://prowl.rockefeller.edu/>.

2.2.3. Determining function of novel molecular marker proteins

In addition to solely isolating stress-regulated genes or proteins to use as molecular markers, it is usually of interest to elucidate the function of any novel gene product. Although outside the main scope of this review, we would like to briefly illustrate how such an activity identification could be achieved.

Database searches and activity confirmation. The first obvious study to perform once the whole or part of the primary structure of a novel stress-related protein has been established, is to perform FASTA or BLAST search in GenBank (<http://www.ncbi.nlm.nih.gov/>) or species-specific databases (e.g. <http://www.arabidopsis.org/> or <http://www.agron.missouri.edu/>). A putative function obtained for the novel polypeptide can sometimes be obtained in this way. For confirmation, activity measurements (for putative enzymes), binding studies (for structural proteins, transcription regulatory proteins etc.), yeast complementation studies (when an appropriate yeast mutant is available), or any other appropriate assay, have to be performed. However, in many cases no functional information can be obtained from database searches. In these cases, the specific function for a protein has to be investigated in each instance. A number of suggestions for methods to determine protein function are briefly described below.

T-DNA knock-out mutant searches. Where the plant species under study is amenable to *Agrobacterium*-mediated transformation, T-DNA mutagenesis may be a useful tool to study gene function. In plants carrying a mutation in the gene of interest, causing either a complete or partial loss of function, a study of the phenotype may give valuable clues

to the role of the gene product. The phenotype under control conditions and under the influence of an abiotic stress may be compared. However, one should realise that the risk that the mutation gives rise to pleiotropic effects is great, especially if a transcription factor is studied. In addition, a lack of observable phenotypical effects does not necessarily mean that the gene of interest is not important for the tolerance towards abiotic stress. Redundancy caused by the expression of a number of other isoproteins could partly fulfill the function of an inactivated member of a gene family. In addition, the severity of the mutation caused by T-DNA integration can vary depending on the location of the insertion with respect to the gene and also on the nature of the sequences carried by the T-DNA. Ideally, a null mutation will be available for study, but often expression is merely reduced and in a few cases increased expression has been observed (Krysan *et al.*, 1999)

To more rapidly obtain a function for a gene in a crop species, particularly those that are not amenable to T-DNA mutagenesis, it is often convenient and productive to examine the equivalent gene in *Arabidopsis thaliana*. A very powerful way to obtain, with a high probability, the desired mutant in *A. thaliana* is to utilise the *Arabidopsis* gene knock-out facility in Madison, Wisconsin (<http://www.biotech.edu/Arabidopsis/>), and to search their over 60,000 T-DNA-tagged lines for a suitable mutant (Krysan *et al.*, 1999). This procedure requires knowledge of the genomic nucleotide sequence of the *Arabidopsis* gene equivalent to the gene of interest in the particular crop species studied (see above). Since the *Arabidopsis* genome sequencing project is nearing completion, a search in the *Arabidopsis* genome database (<http://www.arabidopsis.org/agi.html>) should provide the interested researcher with candidate genes. After having successfully received mutants from the knock-out facility, the *Arabidopsis* phenotype can be studied under both control and stress conditions. Gene function can also be restored by complementation with first a functional copy of the *Arabidopsis* gene and then eventually with the gene obtained from the original crop species.

Two-hybrid system. The two-hybrid system in yeast is a method used to identify *in vivo* protein-protein interactions (Vidal and Legrain, 1999). This technique takes advantage of the domain-based structure of transcription factors. Most transcription factors consist of a DNA-binding domain and an RNA polymerase activating domain. The gene under investigation is cloned into a special vector where it will be expressed as a fusion protein together with a DNA-binding domain (the "bait" vector). Then, a cDNA library is created from the species under investigation where each open reading frame is expressed as fusion protein with a RNA polymerase-activating domain (the "prey" vector). The bait

and the prey are then co-transformed into a yeast strain in which multiple DNA binding sequences have been placed in front of reporter genes (usually *lacZ* and a growth selection marker *LEU2* or *HIS3*). If a protein-protein interaction occurs, the reporter genes are activated. A potential use of this technique in stress molecular marker analysis is to find regulatory proteins of known biosynthetic pathways, for example.

One-hybrid system. The one-hybrid system is a method used to find DNA-binding proteins (Vidal and Legrain, 1999). Prior knowledge of a promoter or enhancer DNA sequence is required. Multiple copies of the DNA sequence in question are cloned in front of reporter genes and inserted into a yeast strain. The new strain is then transformed with a cDNA expression library where the polypeptides are expressed as fusion proteins with an RNA polymerase-activating domain (i.e. the same library that is used in the two hybrid system can be used). If a protein binds to the DNA sequence, the reporter genes will be activated and the protein may then be identified. The one-hybrid system has been successfully used to find bZIP transcription factors that interact with ABA-responsive DNA elements (Kim *et al.*, 1997).

2.2.4. Activity measurements

Even though protein quantification methods, such as those involving the use of polyclonal antibodies, can be successfully used to determine the concentration of a given protein in a tissue, most antibodies do not discriminate between functional and non-functional polypeptides. In cases where, rather than giving rise to an increased flow of information from gene to protein, the cellular response to an environmental factor is manifested as an activation or deactivation of preformed protein molecules (for example through covalent modification), it is necessary to investigate the changes in activity of the enzyme or the binding capacity of the protein of interest. Assays have been developed for a multitude of enzymatic reactions and in the case of novel activities, the approaches differ between reaction types. It is thus beyond the scope of this review to guide the reader in the development of methods for enzymatic and binding activity measurements. Instead, an appropriate standard method for each specific activity should be used.

2.3. Choosing potential markers

Most of the methods described above have been used to identify novel differentially regulated genes or proteins or to follow global gene

expression during and/or after stress. In many cases such studies are not required to obtain a useful molecular marker. For instance, if the marker is to be used just to check whether a particular biochemical pathway is regulated during stress, simply cloning the differentially regulated gene of interest from genomic DNA or cDNA by using PCR may be adequate and will save a considerable amount of time and effort (provided the appropriate sequence information is available). If a gene is known to be regulated in one species, this information can be used either to clone the gene of interest from other species using degenerate primers and PCR, or if the species are closely related, the same marker probe might even be used.

Another point to consider is the specificity of the marker to be used. In some application this might be of critical importance. For instance if a specific marker is to be found for ozone exposure, the same marker should not also be activated by other stresses such as cold or pathogen attack. Most studies concerning stress-regulated genes are performed with one stress only. In the few studies where genes have been tested for responsiveness to different stresses (Brosché and Strid, 1999b; Marrs and Walbot, 1997; Richards *et al.*, 1998; Sävenstrand *et al.*, 2000; Zegzouti *et al.*, 1997), very few genes have been found to be uniquely regulated by one stress only. To find a truly stress specific gene, the use of subtractive hybridisation or suppression subtractive hybridisation is preferred. The driver probe could include mRNA/cDNA from plants exposed to other stresses leaving potential stress-specific cDNAs in the tester probe.

For genes responsive to environmental stress, the general pattern seems to be a regulation of gene expression by several different biotic or abiotic stresses and/or phytohormones. However, with the advent of DNA microarrays it is now possible to simultaneously detect expression levels of thousands of genes. It is probable that a stress-specific "fingerprint" based on the expression profiles of thousand of genes can be found. Many important crop species have EST projects in progress (Richmond and Somerville, 2000) which will permit widespread exploitation of DNA microarrays in the near future.

For UV-B radiation stress a highly specific alternative molecular marker is already available in the form of DNA damage. UV-B radiation specifically interacts with the bases of DNA, forming cyclobutane-pyrimidine dimers and pyrimidine (6-4) pyrimidone photoproducts (Thoma, 1999). For both types of damage, specific monoclonal antibodies are available which have already been successfully used in plants (Nakagawa *et al.*, 1998; Stapleton *et al.*, 1993). These antibodies can be used in a simple ELISA to detect the relative amounts of DNA damage.

3. DETECTION UNDER STRESS

Once a stress-specific molecular marker has been isolated, characterised, and been found to stand the tests aimed at validating its appropriateness as a diagnostic tool, a detection method with which to routinely use the marker has to be chosen. A number of suggestions follows in this section.

3.1. Detection of mRNAs

3.1.1. Northern blotting and slot blotting

One of the most commonly used methods to detect the expression of particular mRNA transcripts is northern blotting. Once a cDNA has been identified by using one of the methods described above, it is initially used to confirm the specific regulation of transcription, since most gene isolation methods can give rise to false positives. Northern blotting is a simple method but requires the mRNA of interest to be present at rather a high concentration (Table 3), i.e. northern blot analysis is not as sensitive as RT-PCR (see below).

Total RNA is loaded on a denaturing agarose gel and separated by electrophoresis. The RNA is transferred to a nylon membrane and cross-linked to the surface by UV-irradiation or baking at high temperature. The membrane is then incubated with the specific, nescence or radioactivity) molecular probe and the mRNA of interest, which hybridises with the probe, can be detected on an X-ray film.

If a large number of samples is to be investigated, slot or dot blotting is an alternative to northern blotting. In this case, total RNA is applied directly to the nylon membrane, either by dotting the sample by hand or using commercially available manifold devices, and the membrane is hybridised as described above. The major advantage with slot blotting is that the screening of several RNA samples can be performed in a short time. However, with slot blotting, as compared to northern blotting, information concerning the size of the hybridising mRNA species is lost. In addition, slot blotting is not a suitable alternative in cases where the probe hybridises to several mRNA species (Sambrook et al., 1989). Thus, prior to using dot or slot blotting, the banding pattern has to be established by northern blotting.

3.1.2. RT-PCR

RT-PCR (reverse transcription – polymerase chain reaction) is the most sensitive technique available to monitor gene expression (Table 3). In theory using this technique it is possible to detect single copy mRNA transcripts. The first step consists of isolation of total RNA or mRNA followed by cDNA synthesis from this RNA, while the second step is the amplification of the cDNA species of interest using specific primers and PCR. An RT-PCR experiment is fast: usually the whole procedure can be performed in one day or less. RT-PCR can also be used to detect differences between closely related sequences using discriminating primers (Dieffenbach *et al.*, 1995). Furthermore, a large number of samples can be processed simultaneously.

The major drawback with RT-PCR is that the method is difficult to control and optimise. Both the cDNA synthesis step and the PCR step may proceed with varied efficiencies in different reactions which leads to differences in the product yield, even between replicates of the same sample (Gause and Adamovicz, 1995). Impurities in the RNA might disturb the enzymatic reactions and traces of genomic DNA can give rise to false results. The latter problem is solved by performing a DNase I treatment of the RNA and by including a "no-RT" control in the absence of reverse transcriptase. The primers may also be designed to comprise the borders of introns leading to differently sized PCR products depending on whether they are amplified from cDNA or genomic DNA. For accurate quantification, the RT-PCR is performed either as a competitive or a relative RT-PCR.

In competitive RT-PCR (Becker-André, 1993; Chang *et al.*, 1993), an *in vitro*-synthesised RNA transcript is added during the RT step. This competitive transcript is then amplified together with the target cDNA during the PCR. The amounts of target and competitive PCR product are then compared. If the amounts are equal it is assumed that the concentration of the target RNA is equal to the concentration of the added competitor. Both the target and the competitor RNA must be in the same concentration range to get accurate quantification. Competitive RT-PCR requires equal amplification efficiency of competitor and target cDNAs and also requires a method to discriminate between the two products.

In relative RT-PCR (Gause and Adamovicz, 1995), a second primer set is added to the PCR, amplifying an internal standard, usually a "house-keeping" gene. The level of target product is then adjusted to the level of the internal standard product. The products must be analysed while the PCR is still in linear amplification for both the target and the internal standard cDNAs. If not, the differences in the amounts between samples will be lost (Gause and Adamovicz, 1995). The choice of internal

Table 3. Comparison between methods for detection of levels of mRNAs in plant tissue

	NORTHERN BLOTTING	RT-PCR	RNase PROTECTION ASSAY	DNA ARRAYS
Estimated sensitivity	Low	High	Medium	Low to medium
Level of optimisation required	Low	High	Medium	Medium
Cost	Low	Medium	Medium	Very high
Determination of mRNA size	Yes	No	No	No
Detection of alternatively spliced transcripts	Yes	No	Yes	No
Distinguish between members of multi-gene families	No	Yes	Yes	No
Number of mRNAs detected simultaneously	1	1	1-6	100-10000
Tolerates partially degraded RNA	No	Yes	Yes	No
Labour intensity	Low	High	Low	High

standard is critical, ideally it should be a gene the expression of which is independent of environmental factors.

The latest development in the area of RT-PCR is the use of "real time" PCR where the formation of product is followed during every cycle using fluorescent probes. Several different pieces of equipment are available on the market. "Real time" PCR solves the problem of maintaining the linear amplification phase and eliminates the post-PCR detection phase.

3.1.3. Ribonuclease protection assay

The basis of the ribonuclease protection assay is a solution hybridisation between a single-stranded radiolabelled anti-sense riboprobe to an RNA sample. After hybridisation, excess unhybridised probe and sample RNA are removed by ribonuclease digestion leaving probe:target RNA hybrids unaffected. These hybrids are precipitated, separated on a denaturing polyacrylamide gel and visualised by autoradiography. The advantages of ribonuclease protection compared to northern blotting are a 10-100 times higher sensitivity, the fact that several different probes may be used in the same hybridisation as long as their PAGE gel mobilities are different, and that the method is able to discriminate between closely related sequences such as differently sized messages emanating from the same gene (Table 3). The RNase protection assay is also tolerant of partially degraded RNA samples. As an example, the RNase protection assay has been used to determine the effect of cadmium stress on splicing of *Bronze2* mRNA transcripts in *Zea mays* (Marrs and Walbot, 1997).

3.1.4. DNA arrays

As described above, large scale EST and cDNA sequencing projects have generated a considerable amount of sequence information. This may be exploited to monitor global changes in gene expression using DNA arrays (Desprez *et al.*, 1998; Ruan *et al.*, 1998; Richmond and Somerville, 2000). The principle underlying this method is hybridisation of a labelled total transcript probe (cDNA or complementary RNA) to a solid matrix carrying DNA sequences representing a large number (more than 1000) of different genes arranged in an ordered manner. Examination of the resulting hybridisation pattern allows the expression level of many genes to be analysed simultaneously. Currently, three different systems are available for performing a DNA array experiment: nylon filters spotted with cDNAs, glass slides spotted with cDNAs and GeneChips (Affymetrix

Inc.) carrying oligonucleotides which have been synthesised directly onto the chip surface.

Nylon membrane-based cDNA arrays (macroarrays) carry between 100 and 1000 PCR-amplified cDNA fragments spotted at high density onto a nylon filter. One membrane is required for each treatment to be compared. RNA samples (isolated from control and stress-exposed plants) are reversely transcribed in the presence of ^{32}P -labelled nucleotides. Each sample is then used to probe a separate membrane and after hybridisation and washing the blots are analysed using a phosphorimager for accurate quantification of the bound radioactivity. Control spots must be included to permit normalisation between membranes, which are used for one treatment only. This simple DNA array method is labour intensive but does not require special equipment, other than a phosphorimager, and so is widely applicable.

A slide DNA microarray is constructed by mechanically spotting DNA fragments (usually PCR products amplified from cDNAs or ESTs) onto glass microscope slides. RNA is isolated from plant material grown under the conditions that are to be compared (i.e. control and stress-exposed) and these two samples are separately reversely transcribed in the presence of two different fluorescent labels. The resulting fluorescent probes are then hybridised to the same slide. After hybridisation and washing the slide is scanned using a fluorescent or confocal microscope, or a commercially available reader. To quantify differential expression, the difference in the intensity of the two bound fluorescent probes is compared for each spot using special computer programs. The robot employed to produce slide microarrays is available from several commercial vendors or can be "homebuilt" (see Ferea and Brown, 1999, for a table of websites for making and using DNA arrays, and Harrington *et al.*, 2000, for a list of resources and software for microarray data analysis). It is also possible to have custom-made microarrays manufactured by commercial laboratories.

GeneChip microarrays available from Affymetrix Inc. (Santa Clara, CA) carry oligonucleotides, representing thousands of genes, synthesised directly onto the chip. The oligonucleotides are synthesised utilising a light-directed chemical process, which combines solid-phase chemical synthesis with photolithographic fabrication techniques used in the semiconductor industry (Harrington *et al.*, 2000; <http://www.affymetrix.com/>). The RNA sample to be analysed is isolated and labelled in a two-step process to produce biotinylated complementary RNA (cRNA) probe. After the hybridisation reaction is complete, biotin-cRNA bound to the GeneChip is stained with a fluorophore conjugated to avidin, and the array is inserted into a scanner where patterns of hybridisation are visualised. To compare different treatments, it is

necessary to perform independent hybridisations with separate chips and then to normalise and analyse the data using special computer software.

Due to the large number of cDNAs or ESTs that can be spotted onto an array, this technique allows the monitoring of expression of thousands of genes at the same time (Table 3). This makes it possible to determine a unique gene expression fingerprint for each stress condition of interest. The major drawback with the technique is the large initial investment needed for equipment and the laborious collection and preparation of cDNAs for manufacturing the arrays.

Plant biology in general and stress research in particular will benefit greatly by the imminent completion of the *Arabidopsis* genome sequencing project. Once this milestone has been achieved, every gene or potential gene in this model plant genome may be identified. This will permit total ORF (open reading frame) microarrays to be constructed, as have already been produced for yeast and some bacteria. Such arrays will be invaluable in describing the changing patterns of transcript levels caused by any stress imaginable. Not only will stress responsive genes be identified but whole networks of genes showing coordinated expression patterns will also be discovered which will significantly enhance understanding of plant stress responses. Findings in the *Arabidopsis* model system will undoubtedly be applicable to most plants and will guide future research into improving crop tolerance to stress.

At the web pages of the *Arabidopsis* Functional Genomics Consortium and in related databases, (http://afgc.stanford.edu/afgc_html/site1.htm), it is possible to find gene expression data from studies on *Arabidopsis* exposed to a variety of different treatments.

3.2. Detection of protein

Although in principle specific proteins can routinely be detected by ordinary one-dimensional polyacrylamide gel electrophoresis and staining, or in some cases two-dimensional electrophoresis which may be followed by mass spectrometry analysis, immunological detection methods are favoured due to their specificity and ability to detect minute amounts of protein.

In addition, from a diagnostic point of view, more laborious immunological methods, such as immunohistochemistry or immunoblotting, are less favoured than high throughput techniques such as ELISA. All these immunological methods rely on the availability of antibodies towards the protein of interest. In addition, protein extracts of the plant tissue to be assayed have to be obtained. In this context, it is of importance to have sufficient knowledge about the localisation of the

protein in such extracts. For instance, is the protein soluble during extraction, or is it co-precipitating with other cellular components, e.g. membrane fractions, during centrifugation? In the latter case, inclusion of solubilising agents, such as detergents, which should not interfere with the immunological reaction, is commonplace. Modifications of methods have to be considered in each specific case and will not be discussed further in this review.

4. CONCLUSIONS

In this review we have discussed how molecular markers for anthropogenic environmental stresses may be identified and how variations in their expression can be monitored. The various procedures to identify differentially expressed plant genes can potentially be used to identify novel stress-responsive genes in any crop plant (provided that RNA can be extracted from the tissue under study). However, the reader must decide if their particular problem requires the identification of novel genes. In situations where it is necessary only to determine whether or not a plant is experiencing stress, well characterised marker genes may be used. These include, for example, genes encoding enzymes of the phenylpropanoid/flavonoid biosynthetic pathway (Dixon and Paiva, 1995), pathogen response proteins (Datta and Muthukrishnan, 1999), cell wall proteins (Showalter, 1993) and free radical scavenging enzymes (Nocter and Foyer, 1998).

It should also be recognised that at present very few stress-specific marker genes have been identified. Due to the common signalling pathways employed by plants in response to different stresses, largely the same set of genes is induced or repressed. With the imminent completion of the *Arabidopsis* genome project and the production of total ORF micro-arrays it is likely that more stress-specific markers will be identified. Micro-arrays which carry DNA sequences representing the total complement of *Arabidopsis* genes will permit the complete pattern of gene expression to be characterised for any stress or combination of stresses imaginable. Data from this model plant may be directly applicable to related species (eg. cabbage, radish) or can direct further research to identify more specific marker genes in other crops.

It can be envisaged that early detection of plant stress by monitoring the expression of a number of key marker genes will become an essential tool for large-scale arable farming in the future. Ideally a panel of genes showing characteristic patterns of expression under a spectrum of stress conditions would be the basis of a rapid diagnostic test. This test may either be performed on samples of tissue taken from randomly selected plants or could rely on biosensor plants distributed

amongst the crop. In the former situation, levels of mRNA transcripts for each marker gene would be quantified, while in the latter, the activity of reporter enzymes expressed under control of specific stress regulated promoters would be assayed. These tests should be sufficiently rapid to allow measures to combat a particular stress to be implemented within minutes to hours of taking the tissue samples. This need for a quick response would appear to necessitate a test reliant on PCR when measuring mRNA transcript levels or a simplified reporter enzyme assay when examining promoter activity. In both cases the test needs to be sufficiently simple and robust for it to be performed literally "in the field".

5. References

- Andersen, J.S. and Mann, M. (2000) Functional genomics by mass spectrometry, *FEBS Letters* **480**, 25-31.
- Axel, R., Feigelson, P. and Schutz, G. (1976) Analysis of the complexity and diversity of mRNA from chicken liver and oviduct, *Cell* **7**, 247-254.
- Bachem, C.W.B., Oomen, R.J.F.J. and Visser, R.G.F. (1998) Transcript imaging with cDNA-AFLP: A step-by-step protocol, *Plant Molecular Biology Reporter* **16**, 157-173.
- Bachem, C.W.B., van der Hoeven, R.S., de Bruijn, S.M., Vreugdenhil, D., Zabeau, M., and Visser, R.G.F. (1996) Visualization of differential gene expression using a novel method of RNA fingerprinting based on AFLP: analysis of gene expression during potato tuber development, *Plant Journal* **9**, 745-753.
- Bailey-Serres, J. (1999) Selective translation of cytoplasmic mRNAs in plants, *Trends in Plant Science* **4**, 142-148.
- Balzer, H.-J. and Bäumlein, H. (1994) An improved gene expression screen, *Nucleic Acids Research* **22**, 2853-2854.
- Bar-Peled, M., Bassham, D.C. and Raikhel, N.V. (1996) Transport of proteins in eukaryotic cells: more questions ahead, *Plant Molecular Biology* **32**, 223-249.
- Becker-André, M. (1993) Absolute levels of mRNA by polymerase chain reaction-aided transcript titration assay, *Methods in Enzymology* **218**, 420-445.
- Bertioli, D.J., Schlichter, U.H.A., Adams, M.J., Burrows, P.R., Steinbiß, H.-H. and Antoniw, J.F. (1995) An analysis of differential display shows a strong bias towards high copy number mRNAs, *Nucleic Acids Research* **23**, 4520-4523.
- Bishop, J.O., Morton, J.G., Rosbash, M. and Richardson, M. (1974) Three abundant classes in HeLa cell messenger RNA, *Nature* **250**, 199-204.
- Brosché, M. and Strid, Å. (1999a) The mRNA-binding ribosomal protein S26 as a molecular marker in plants: molecular cloning, sequencing and differential gene expression during environmental stress, *Biochimica et Biophysica Acta* **1445**, 342-344.

Brosché, M. and Strid, Å. (1999b) Cloning, expression, and molecular characterization of a small pea gene family regulated by low levels of ultraviolet B radiation and other stresses, *Plant Physiology* **121**, 479-487.

Brosché, M., Fant, C., Bergkvist, S.W., Strid, H., Svensk, A., Olsson, O. and Strid, Å. (1999) Molecular markers for UV-B stress in plants: alteration of the expression of four classes of genes in *Pisum sativum* and the formation of high molecular mass RNA adducts, *Biochimica et Biophysica Acta* **1447**, 185-198.

Buchanan-Wollaston, V. and Ainsworth, C. (1997) Leaf senescence in *Brassica napus*: cloning of senescence related genes by subtractive hybridisation, *Plant Molecular Biology* **33**, 821-834.

Caldwell, C.R. (1998) Effect of elevated manganese on the ultraviolet- and blue light-absorbing compounds of cucumber cotyledons and leaf tissue, *Journal of Plant Nutrition* **21**, 435-445.

Caldwell, M.M., Teramura, A.H., Tevini, M., Bornman, J.F., Björn, L.O. and Kulandaivelu, G. (1994) Effects of Increased solar ultraviolet radiation on terrestrial plants, in *Environmental effects of ozone depletion: 1994 assessment*, United Nations Environment Programme, Nairobi, chapt. 4.

Chang, P.F., Narasimhan, M.L., Hasegawa, P.M. and Bressan, R.A. (1993) Quantitative mRNA-PCR for expression analysis of low abundance transcripts, *Plant Molecular Biology Reporter* **11**, 237-248.

Chen, J.-J., Rowley, J.D. and Wang, S.M. (2000) Generation of longer cDNA fragments from serial analysis of gene expression tags for gene identification, *Proceedings of the National Academy of Sciences of the United States of America* **97**, 349-353.

Choi, S.-Y., Baek, E.-M. and Lee, S.Y. (1995) A cDNA differentially expressed by cadmium stress in *Arabidopsis*, *Plant Physiology* **108**, 849.

Clark, S. (1993) Protein methylation, *Current Opinion in Cell Biology* **5**, 977-983.

Coligan, J.E., Kruisbeek, A.M., Margulies, D.H., Shevach, E.M. and Strober, W. eds. (1997) *Current Protocols in Immunology*, John Wiley & Sons, New York.

Cooke, R., Raynal, M., Laudié, M., Grellet, R., Delseny, M., Morris, P.-C., Guerrier, D., Giraudat, J., Quigley, F., Clabault, G., Li, Y.-F., Mache, R., Krivitzky, M., Gy, I.J.-J., Kreis, M., Lechamy, A., Parmentier, Y., Marbach, J., Fleck, J., Clément, B., Philipps, G., Hervé, C., Bardet, C., Tremousaygue, D., Lescure, B., Lacomme, C., Roby, D., Jourjon, M.-F., Chabrier, P., Charpentreau, J.-L., Desprez, T., Amselem, J., Chiapello, H. and Höfte, H. (1996) Further progress towards a catalogue of all *Arabidopsis* genes: analysis of a set of 5,000 non-redundant ESTs, *The Plant Journal* **9**, 101-124.

Courchesne, P.L. and Patterson, S.D. (1999) Identification of proteins by matrix-assisted laser desorption/ionization mass spectrometry using peptide and fragment ion masses, in A.J. Link (ed.), *2-D Proteome Analysis Protocols*, Humana Press Inc., New Jersey, pp. 487-511.

- Cruz-Ortega, R. Cushman, J.C. and Ownby, J.D. (1995) Nucleotide sequence of cDNA for a 1,3-beta glucanase associated with aluminum toxicity in wheat roots (Accession no. U30323) (PGR95-073), *Plant Physiology* **109**, 722.
- Datson, N.A., van der Perk-de Jong, J., van den Berg, M.P., de Kloet, E.R. and Vreugdenhil, E. (1999) MicroSAGE: a modified procedure for serial analysis of gene expression in limited amounts of tissue, *Nucleic Acids Research* **27**, 1300-1307.
- Datta, S.K. and Muthukrishnan, S. eds. (1999) *Pathogenesis-Related Proteins in Plants*, CRC press, Washington, D.C.
- Desprez, T., Amselem, J., Caboche, M. and Höfte, H. (1998) Differential gene expression in *Arabidopsis* monitored using DNA arrays, *The Plant Journal* **14**, 643-652.
- Diatchenko, L., Lau, Y.-F.C., Campbell, A.P., Chenchik, A., Moqadam, F., Huang, B. Lukyanov, S., Lukyanov, K., Gurskaya, N., Sverdlov, E.D. and Sierbert, P.D. (1996) Suppression subtractive hybridization: A method for generating differentially regulated or tissue-specific cDNA probes and libraries. *Proceedings of the National Academy of Sciences of the United States of America* **93**, 6025-6030.
- Diatchenko, L., Lukyanov, S., Lau, Y.F. and Siebert, P.D. (1999) Suppression subtractive hybridization: a versatile method for identifying differentially expressed genes, *Methods in Enzymology* **303**, 349-80.
- Didierjean, L., Frendo, P., Nasser, W., Genot, G., Marivet, J. and Burkard, G. (1996) Heavy-metal-responsive genes in maize: identification and comparison of their expression upon various forms of abiotic stress, *Planta* **199**, 1-8.
- Dixon, R.A. and Paiva, N.L (1995) Stress-induced phenylpropanoid metabolism, *The Plant Cell* **7**, 1085-1097.
- Dupree, P. and Sherrier, D.J. (1998) The plant golgi apparatus, *Biochimica et Biophysica Acta* **1404**, 259-270.
- Eckey-Kaltenbach, H., Kiefer, E., Grosskopf, E., Ernst, D. and Sandermann Jr., H. (1997) Differential transcript induction of parsley pathogenesis-related proteins and of a small heat shock protein by ozone and heat shock, *Plant Molecular Biology* **33**, 343-350.
- Edelman, A.M., Blumenthal, D.K. and Krebs, E.G. (1987) Protein serine/threonine kinases, *Annual Review of Biochemistry* **56**, 567-613.
- Eng, J.K., McCormack, A.L. and Yates, J.R. III (1994) An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database, *Journal of the American Society for Mass Spectrometry* **5**, 976-989.
- Ezaki, B., Yamamoto, Y. and Matsumoto, H. (1995) Cloning and sequencing of the cDNAs induced by aluminium treatment and P_i starvation in cultured tobacco cells, *Physiologia Plantarum* **93**, 11-18.
- Ezaki, B., Tsugita, S. and Matsumoto, H. (1996) Expression of a moderately anionic peroxidase is induced by aluminium treatment in tobacco cells: possible involvement of peroxidase isozymes in aluminum ion stress, *Physiologia Plantarum* **96**, 21-28.

Ezaki, B., Koyanagi, M., Gardner, R.C. and Matsumoto, H. (1997) Nucleotide sequence of a cDNA for GDP dissociation inhibitor (GDI) which is induced by aluminum (Al) ion stress in tobacco cell culture (Accession No. AF012823) (PGR97-133), *Plant Physiology* **115**, 314.

Ferea, T.L. and Brown, P.O. (1999) Observing the living genome, *Current Opinion in Genetics and Development* **9**, 715-722.

Fichmann, J. and Westermeier, R. (1999) 2-D protein gel electrophoresis, in A.J. Link (ed.), *2-D Proteome Analysis Protocols*, Humana Press Inc., New Jersey, pp. 1-7.

Fukuda, T., Kido, A., Kajino, K., Tsutsumi, M., Miyauchi, Y., Tsujiuchi, T., Konishi, Y. and Hino, O. (1999) Cloning of differentially expressed genes in highly and low metastatic rat osteosarcomas by a modified cDNA-AFLP method, *Biochemical and Biophysical Research Communications* **261**, 35-40.

Furuyama, T. and Dzelzkalns, V.A. (1999) A novel calcium-binding protein is expressed in *Brassica* pistils and anthers late in flower development, *Plant Molecular Biology* **39**, 729-737.

Gause, W.C. and Adamovicz, J. (1995) Use of PCR to quantitate relative differences in gene expression, in C.W. Dieffenbach and G.S. Dveksler (eds.), *PCR primer: A laboratory manual*, Cold Spring Harbor Laboratory Press, New York, pp. 293-311.

Gutiérrez, R.A., MacIntosh, G.C. and Green, P.J. (1999) Current perspectives on mRNA stability in plants: multiple levels and mechanisms of control, *Trends in Plant Science* **4**, 429-438.

Hare, P.D., Du Pleissis, S., Cress, W.A. and Van Staden, J. (1996) Stress-induced changes in plant gene expression, *South African Journal of Science* **92**, 431-439.

Harrington, C.A., Rosenow, C. and Retief, J. (2000) Monitoring gene expression using DNA microarrays, *Current Opinion in Microbiology* **3**, 285-291.

Ivashuta, S., Imai, R., Uchiyama, K. and Gau, M. (1999) The coupling of differential display and AFLP approaches for nonradioactive mRNA finger printing, *Molecular Biotechnology* **12**, 137-141.

Jensen, P.K., Pasa-Tolic, L., Anderson, G.A., Horner, J.A., Lipton, M.S., Bruce, J.E. and Smith, R.D. (1999) Probing proteomes using capillary isoelectric focusing-electrospray ionization Fourier transform ion cyclotron resonance mass spectrometry, *Analytical Chemistry* **71**, 2076-2084.

Jung, R., Scott, M.P., Nam, Y.W., Beaman, T.W., Bassuner, R., Saalbach, I., Muntz, K. and Nielsen, N.C. (1998) The role of proteolysis in the processing and assembly of 11S seed globulins, *The Plant Cell* **10**, 343-357.

Kiiskinen, M., Korhonen, M. and Kangasjärvi, J. (1997) Isolation and characterization of cDNA for a plant mitochondrial phosphate translocator (*Mpt1*): ozone stress induces *Mpt1* mRNA accumulation in birch (*Betula pendula* Roth), *Plant Molecular Biology* **35**, 271-279.

- Kim, S.Y., Chung, H.-J. and Thomas, T.L. (1997) Isolation of a novel class of bZIP transcription factors that interact with ABA-responsive and embryo-specification elements in the *Dc3* promoter using a modified yeast one-hybrid system, *The Plant Journal* **11**, 1237-1251.
- Kim, J.Y., Chung, Y.S., Ok, S.H., Lee, S.G., Chung, W.I., Kim, I.Y. and Shin, J.S. (1999) Characterization of the full-length sequences of phospholipase A₂ induced during flower development, *Biochimica et Biophysica Acta* **1489**, 389-392.
- Krutchinsky, A.N., Zhang, W. and Chait B.T. (2000) Rapidly switchable matrix-assisted laser desorption/ionisation and electrospray quadrupole-time-of-flight mass spectrometry for protein identification, *Journal of the American Society for Mass Spectrometry* **11**, 493-504.
- Krysan, P.J., Young, J.C. and Sussman, M.R. (1999) T-DNA as an insertional mutagen in *Arabidopsis*, *The Plant Cell* **11**, 2283-2290.
- Kuc, J. (1995). Phytoalexins, stress metabolism and disease resistance in plants, *Annual Review of Phytopathology* **33**, 275-297.
- Lers, A., Burd, S., Lomaniec, E., Droby, S. and Chalutz, E. (1998) The expression of a grapefruit gene encoding an isoflavone reductase-like protein is induced in response to UV irradiation, *Plant Molecular Biology* **36**, 847-856.
- Li, W., Hendrickson, C.L., Emmett, M.R. and Marshall, A.G. (1999) Identification of intact proteins in mixtures by alternated capillary liquid chromatography electrospray ionisation and LC ESI infrared multiphoton dissociation Fourier transform ion cyclotron resonance mass spectrometry, *Analytical Chemistry* **71**, 4397-4402.
- Liang, P. and Pardee, A.B. (1992) Differential display of eukaryotic messenger RNA by means of the polymerase chain reaction, *Science* **257**, 967-971.
- Link, A.J., Eng, J., Schieltz, D.M., Carmack, E., Mize, G.J. Morris, D.R., Garvik, B.M., Yates, J.R. III (1999) Direct analysis of protein complexes using mass spectrometry, *Nature Biotechnology* **17**, 676-682.
- Lodl, P. (1994) Histone acetylation: facts and questions, *Chromosoma* **103**, 441-449.
- Lorkovic, Z.J., Wieczorek-Kirk, D.A., Lambermon, M.H.L. and Filipowicz, W. (2000) Pre-mRNA splicing in higher plants, *Trends in Plants Science* **5**, 160-167.
- Macas, J., Lambert, G.M., Dolezel, D., and Galbraith, D.W. (1998). NEST (Nuclear Expressed Sequence Tag) analysis: a novel means to study transcription through amplification of nuclear RNA. *Cytometry* **33**, 460-468.
- Mann, M. (1999) Quantitative proteomics?, *Nature Biotechnology* **17**, 954-955.
- Mann, M. and Wilm, M. (1994) Error tolerant identification of peptides in sequence databases by peptide sequence tags, *Analytical Chemistry* **66**, 4390-4399.
- Marrs, K.A. and Walbot, V. (1997) Expression and RNA splicing of the maize glutathione S-transferase bronze2 gene is regulated by cadmium and other stresses, *Plant physiology* **113**, 93-102.

Matsumura, H., Nirasawa, S. and Terauchi R. (1999) Technical advance: transcript profiling in rice (*Oryza sativa* L.) seedlings using serial analysis of gene expression, *The Plant Journal* **20**, 719-726.

Money, T., Reader, S., Qu, L.J., Dunford, R.P., and Moore, G. (1996) AFLP-based mRNA fingerprinting, *Nucleic Acids Res.* **24**, 2616-2617.

Nakagawa, A., Kobayashi, N., Muramatsu, T., Yamashina, Y., Shirai, T., Hashimoto, M.W., Ikenaga, M. and Mori, T. (1998) Three-dimensional visualization of ultraviolet-induced DNA damage and its repair in human cell nuclei, *Journal of Investigative Dermatology* **110**, 143-148.

Nocter, G. and Foyer, C.H. (1998) Ascorbate and glutathione: keeping active oxygen under control, *Annual Review of Plant Physiology and Plant Molecular Biology* **49**, 249-279.

O'Farrell, M. (1995) ADP ribosylation reactions in plants, *Biochimie* **77**, 486-491.

Ouvrard, O., Cellier, F., Ferrare, K., Tusch, D., Lamaze, T., Dupuis, J.-M. and Casse-Delbart, F. (1996) Identification and expression of water stress- and abscisic acid-regulated genes in a drought-tolerant sunflower genotype, *Plant Molecular Biology* **31**, 819-829.

Oxford dictionary of biochemistry and molecular biology (1997) Smith, A.D., Datta, S.P., Howard Smith, G., Campbell, P.N., Bentley, R. and McKenzie, H.A. (eds.) Oxford University Press, Oxford, U.K.

Pandey, A. and Mann, M. (2000) Proteomics to study genes and genomes, *Nature* **405**, 837-846.

Pirrung, M.C. (1999) Histidine kinases and two-component signal transduction systems, *Chemistry and Biology* **6**, R167-175.

Pleißner, K.-P., Hoffmann, F., Kriegel, K., Wenk, C., Wegner, S., Sahlström, A., Oswald, H., Alt, H. and Fleck, E. (1999) New algorithmic approaches to protein spot detection and pattern matching in two-dimensional electrophoresis gel databases, *Electrophoresis* **20**, 755-765.

Ragland, M. and Soliman, K.M. (1997) Sali5-4a (Accession No. U64866) and Sali3-2 (Accession No. U89693), two genes induced by aluminum in soybean roots (PGR97-071), *Plant Physiology* **114**, 395.

Reid, W.V. and Miller, K.R. (1989). Keeping Options Alive: The Scientific Basis for Conserving Biological Diversity, World Resources Institute, Washington DC.

Richards, K.D. and Gardner, R.C. (1995) pEARL11 (L40380) an *Arabidopsis* member of a conserved gene family, *Plant Physiology* **109**, 1497.

Richards, K.D., Snowden, K.C. and Gardner, R.C. (1994) *wali6* and *wali7*: Genes induced by aluminum in wheat (*Triticum aestivum* L.) roots, *Plant Physiology* **105**, 1455-1456.

- Richards, K.D., Donaldson, S. and Gardner, R.C. (1995) Nucleotide sequence of pEARLI 4 (Accession No. L43081) from *Arabidopsis* (PGR95-098), *Plant Physiology* **109**, 1497.
- Richards, K.D., Schott, E.J., Sharma, Y.K., Davis, K.R. and Gardner, R.C. (1998) Aluminum induces oxidative stress genes in *Arabidopsis thaliana*, *Plant Physiology* **116**, 409-418.
- Richmond, T. and Somerville, S. (2000) Chasing the dream: plant EST microarrays, *Current Opinion in Plant Biology* **3**, 108-116.
- Robertson, L.D., Singh, K.B., Erskine, W., El Moneim, A.M.A. (1996) Useful genetic diversity in germplasm collections of food and forage legumes from West Asia and North Africa, *Genetic Resources and Crop Evolution* **43**, 447-460.
- Rodriguez-Concepcion, M., Yalovsky, S. and Gruissem, W. (1999) Protein prenylation in plants: old friends and new targets, *Plant Molecular Biology* **39**, 865-870.
- Rounsley, S.D., Glodek, A., Sutton, G., Adams, M.D., Somerville, C.R., Venter, J.C. and Kerlavage, A.R. (1996) The construction of *Arabidopsis* expressed sequence tag assemblies, *Plant Physiology* **112**, 1177-1183.
- Ruan, Y., Gilmore, J. and Conner, T. (1998) Towards *Arabidopsis* genome analysis: monitoring expression profiles of 1400 genes using cDNA arrays, *The Plant Journal* **15**, 821-833.
- Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) *Molecular Cloning: A laboratory Manual* (2nd Edition). Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.
- Sagerström, C.G., Sun, B.I. and Sive, H.L. (1997) Subtractive cloning: Past, Present and Future, *Annual Review of Biochemistry* **66**, 751-783.
- Sasaki, T., Song, J., Koga-Ban, Y., Matsui, E., Fang, F., Higo, H., Nagasaki, H., Hori, M., Miya, M., Murayama-Kayano, E., Takiguchi, T., Takasuga, A., Niki, T., Ishimaru, K., Ikeda, H., Yamamoto, Y., Mukai, Y., Ohta, I., Mihadera, N., Havukkala, I. and Minobe, Y. (1994) Toward cataloguing all rice genes: large-scale sequencing of randomly chosen rice cDNAs from a callus cDNA library, *The Plant Journal* **6**, 615-624.
- Shevchenko, A., Wilm, M., Vorm, O. and Mann, M. (1996) Mass spectrometric sequencing of proteins from silver-stained polyacrylamide gels, *Analytical Chemistry* **68**, 850-858.
- Sävenstrand, H., Brosché, M., Ängelsten, M. and Strid, Å. (2000) Molecular markers for ozone stress isolated by suppression subtractive hybridisation: specificity of gene expression and identification of a novel stress-regulated gene, *Plant Cell and Environment* **23**, 689-700.
- Showalter, A.M. (1993) Structure and function of plant cell wall proteins, *The Plant Cell* **5**, 9-23.
- Sharma, Y.K. and Davis, K.R. (1995) Isolation of a novel *Arabidopsis* ozone-induced cDNA by differential display, *Plant Molecular Biology* **29**, 91-98.

- Singh, K.B., Malhorta, R.S., Halila, M.H., Knights, E.J. and Verma, M.M. (1993) Current status and future strategy in breeding chickpea for resistance to biotic and abiotic stresses, *Euphytica* **73**, 137-149.
- Snowden, K.C. and Gardner, R.C. (1993) Five genes induced by aluminum in wheat (*Triticum aestivum* L.) roots, *Plant Physiology* **103**, 855-861.
- Stapleton, A.E., Mori, T. and Walbot, V. (1993) A simple and sensitive antibody-based method to measure UV-induced DNA damage in *Zea mays*, *Plant Molecular Biology Reporter* **11**, 230-236.
- Thoma, F. (1999) Light and dark in chromatin repair: repair of UV-induced DNA lesions by photolyase and nucleotide excision repair, *The EMBO Journal* **18**, 6585-6598.
- Toguri, T., Umemoto, N., Kobayashi, O. and Ohtani, T. (1993) Activation of anthocyanin synthesis genes by white light in eggplant hypocotyl tissues, and identification of an inducible P-450 cDNA, *Plant Molecular Biology* **23**, 933-946.
- Velculescu, V.E., Zhang, L., Vogelstein, B., and Kinzler, K.W. (1995) Serial analysis of gene expression, *Science* **270**, 484-487.
- Vidal, M. and Legrain, P. (1999) Yeast forward and reverse "n"-hybrid systems, *Nucleic Acids Research* **27**, 919-929.
- Wan, J.S., Sharp, S.J., Poirier, G.M.-C., Wagaman, P.C., Chambers, J., Pyati, J., Hom, Y.-L., Galindo, J.E., Huver, A., Peterson, P.A., Jackson, M.R. and Erlander, M.G. (1996) Cloning differentially expressed mRNAs, *Nature Biotechnology* **14**, 1685-1691.
- Wang, Z. and Brown, D.D. (1991) A gene expression screen, *Proceedings of the National Academy of Sciences of the United States of America* **88**, 11505-11509.
- Wegener, A., Gimbel, W., Werner, T., Hani, J., Ernst, D. and Sandermann Jr., H. (1997) Molecular cloning of ozone-inducible protein from *Pinus sylvestris* L. with high sequence similarity to vertebrate 3-hydroxy-3-methylglutaryl-CoA-synthase, *Biochimica et Biophysica Acta* **1350**, 247-252.
- Welsh, J., Chada, K., Dalal, S.S., Chang, R., Ralph, D. and McClelland M.M. (1992) Arbitrary primed PCR fingerprinting of RNA, *Nucleic Acids Research* **20**, 4965-4970.
- Winicov, I. (2000) Molecular markers and abiotic stresses, in S.M.Jain, D.S. Brar and B.S. Ahloowalia (eds.), *Molecular techniques in crop improvement*, Kluwer Academic Publishers, Dordrecht, pp. XXX-YYY.
- Yamamoto, K. and Sasaki, T. (1997) Large-scale EST sequencing in rice, *Plant Molecular Biology* **35**, 135-44.
- Zegzouti, H., Jones, B., Marty, C., Lelièvre, J.-M., Latché, A., Pech, J.-C. and Bouzayen, M. (1997) ER5, a tomato cDNA encoding an ethylene-responsive LEA-like protein: characterization and expression in response to drought, ABA and wounding. *Plant Molecular Biology* **35**, 847-854.

Zhang, W. and Chait, B.T. (2000) ProFound-An expert system for protein identification using mass spectrometric peptide mapping information, *Analytical Chemistry* **72**, 2482-2489.

RANDOM INSERTIONAL MUTAGENESIS IN *ARABIDOPSIS*

Takuyo Ito¹ and Kazuo Shinozaki^{1, 2*}

¹ Laboratory of Plant Molecular Biology, RIKEN Tsukuba Institute, 3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan. ² Plant Mutation Exploration Team, Plant Functional Genomics Research Group, RIKEN Genomics Science Center, 3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan

1.1. INTRODUCTION

Arabidopsis thaliana is now recognized as a model organism for research in plant biology, especially genetics and molecular biology, and is used for the functional analyses of various genes. However, accumulation of genomic and expressed sequence tag (EST) DNA sequence data (Lin *et al.*, 1999; Mayer *et al.*, 1999; Hofte *et al.*, 1993; Newman *et al.*, 1994) revealed a large number of *Arabidopsis* genes remaining functionally uncharacterized. After determination of the *Arabidopsis* whole genome sequence, we will obtain all the *Arabidopsis* genes, but many of the genes will be only annotated as 'hypothetical' or 'putative'. This means that a reverse-genetical approach will remain important to characterize gene functions. One of the best strategies for reverse genetics is based on insertional mutagenesis. The most effective method for insertional mutagenesis is targeted gene disruption. In budding yeast (*Saccharomyces cerevisiae*), this technique is commonly used because of the high frequency of homologous recombination, and the systematic approach to the discovery of gene function is progressing (Oliver, 1996). However, in other organisms including higher plants, this method is still laborious and time consuming, although a few *Arabidopsis* genes have been successfully

disrupted by this technique (Miao and Lam, 1995; Kempin *et al.*, 1997). As an alternative method, the random insertional mutagenesis approach using a transposon or T-DNA [an element of Ti (tumor inducing)-plasmid of *Agrobacterium*] as a mutagen offers a viable method for obtaining insertion mutants for genes of interest. The inserted foreign DNA not only introduces a mutation but also 'tags' the responsible gene. Since this strategy depends on random insertion in the genome, a huge number of transposon or T-DNA insertions has to be prepared for saturated mutagenesis. After construction of this huge insertion library, an effective screening method for isolation of genes of interest has to be developed. These two steps are critical for random insertional mutagenesis.

1.2. INSERTIONAL MUTAGENS

1.2.1. TRANSPOSON MUTAGENESIS

There are no well-characterized endogenous transposons in *Arabidopsis*, although a number of mobile elements have been identified including *Ta* and *Athila* retroelements (Konieczny *et al.*, 1991; Pelissier *et al.*, 1995), and *Tat1* and *Tag1* transposons (Peleman *et al.*, 1991; Tsay *et al.*, 1993). Therefore, extensively characterized maize transposons such as *Activator/Dissociation* (*Ac/Ds*) and *Suppressor-mutator/defective Suppressor-mutator* (*Spm/dSpm*) [identical to *Enhancer/Inhibitor* (*En/I*)] elements have been applied to *Arabidopsis*. These maize elements in a T-DNA vector have been introduced into the heterologous *Arabidopsis* genome by *Agrobacterium*-mediated transformation. At the beginning, the transposition frequency in *Arabidopsis* was too low to be useful. Over the past ten years, the difficulty has been overcome. One of the major improvements is the development of a two-component system. In this system, non-autonomous elements, *Ds* and *dSpm* (=I), are used as insertional mutagens. These elements can be mobilized by *trans*-active *Ac* and *Spm* (=En) transposase genes, respectively. This system has an advantage that non-autonomous elements become stable after removing the transposase genes by segregation, and that a low frequency of excision can be overcome by constructing fusion of the transposase gene to strong

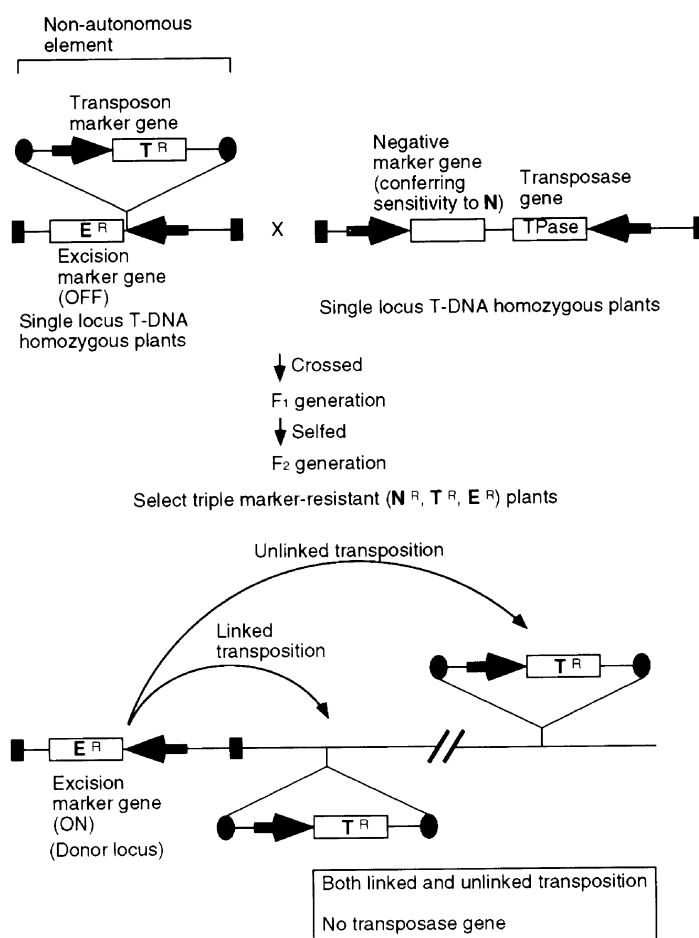


Figure 1.1 Construction and method for local transposition using two-component system

Arrow indicates promoter. Closed circle indicates a border of a non-autonomous element.

Closed rectangle indicates a border of T-DNA. The negative marker gene confers sensitivity to reagent N . The excision and transposon marker genes confer resistance to reagent E and T , respectively. Transformation marker genes necessary for *Agrobacterium*-mediated transformation are omitted.

promoters such as cauliflower mosaic virus (CaMV) 35S promoter (Grevelding *et al.*, 1992).

The most pronounced feature is that transposons tend to transpose to a linked site. Two groups (Fedoroff and Smith, 1993; Machida *et al.*, 1997)

have utilized this feature and developed local transposition systems (*Figure 1.1*). They adopted the *Ac/Ds* two-component system in which non-autonomous *Ds* can be mobilized by a separate *Ac* transposase gene. Two selection markers are used to isolate transposed lines. An excision marker detects an excision event from a start locus, and a transposon marker detects a reinsertion event into the genome. *Ds* is within the excision marker gene, and *Ds* itself contains the transposon marker gene. This engineered construct in a T-DNA vector is introduced into *Arabidopsis*. The other transgenic *Arabidopsis* contains an *Ac* transposase gene controlled by CaMV 35S promoter flanked by a negative marker gene. The negative marker gene confers sensitivity to a reagent, and is used to isolate plants without a transposase gene. By crossing these lines, *Ds* moves during the F1 hemizygous generation. Among F2 plants, transposed lines without the transposase gene are isolated as triple marker-resistant plants (excision, transposon and negative markers). This system has an advantage for local and targeted mutagenesis; 20% of transpositions occurred in the 1-Mb region covering the donor locus (Ito *et al.*, 1999). However, for mutagenesis of the whole genome, many mapped starter lines have to be prepared.

To overcome this disadvantage of the linked transposition system, two groups developed another two-component transposition system (*Ac/Ds*: Sundaresan *et al.*, 1995; *Spm/dSpm*: Tissier *et al.*, 1999), in which closely linked transpositions are discarded, and only unlinked transpositions spread over the genome are obtained (*Figure 1.2*). Since an immobilized negative marker gene flanks the transposable element, only unlinked transpositions are rescued under the negative marker selection. This system has an advantage that a few starter lines are sufficient to obtain insertions spread over the genome without insertion 'hot spots'.

The transposase gene fused to the strong and constitutive CaMV 35S promoter is now commonly used as transposase sources in the *Arabidopsis* two-component systems. However, the disadvantage in using 35S promoter is that the promoter causes successive transpositions that leave footprints, and that it is necessary to introduce the transposase gene back into the

insertion mutant to isolate revertants. For these reasons, inducible transposase activity is preferable. For example, the frequency of transposition of the snapdragon (*Antirrhinum majus*) transposon *Tam3* can be increased by exposing plants to a low temperature, although the induction mechanism is unknown. The excision inducibility of *Tam3* makes it convenient to use for mutagenesis and lineage studies (Carpenter and Coen, 1991). In the case of the *Arabidopsis* two-component system, a heat shock promoter::transposase fusion gene (Balcells *et al.*, 1994) or a transposase::glucocorticoid receptor fusion gene encoding a fusion protein might solve these problems. The glucocorticoid fusion system has been reported to confer glucocorticoid-inducible activity to the fused nuclear protein in higher plants (Aoyama *et al.*, 1995; Simon *et al.*, 1996).

1.2.2. T-DNA MUTAGENESIS

T-DNA, as well as transposons, is an efficient mutagen for insertional mutagenesis in *Arabidopsis*. Advantage of T-DNA mutagenesis, compared with transposons, is that the inserted T-DNA will not mobilize after integration into the genome, and is stable through multiple generations. Another advantage is that T-DNA integrates over the genome, although integration is not completely random. A disadvantage is that T-DNA insertions are often complex. Concatemers arranged as direct or inverted repeats, chromosomal rearrangement, or truncated integration, sometimes make molecular analysis difficult (Gheysen *et al.*, 1990; Castle *et al.*, 1993; Nacry *et al.*, 1998). Recent improvements in *Agrobacterium*-mediated transformation techniques have made the T-DNA a more valuable mutagen. Especially, the seed transformation method (Feldmann and Marks, 1987) and *in planta* transformation method (Bechtold *et al.*, 1993; Clough and Bent, 1998) enable the production of a huge number of insertional mutations necessary for genome-wide saturation of *Arabidopsis* genes. By using these methods, a relatively low number of T-DNA inserts was integrated (1.4 inserts per genome; Feldmann, 1991). One of the large-scale T-DNA insertion libraries has been described by Krysan *et al.* (1999).

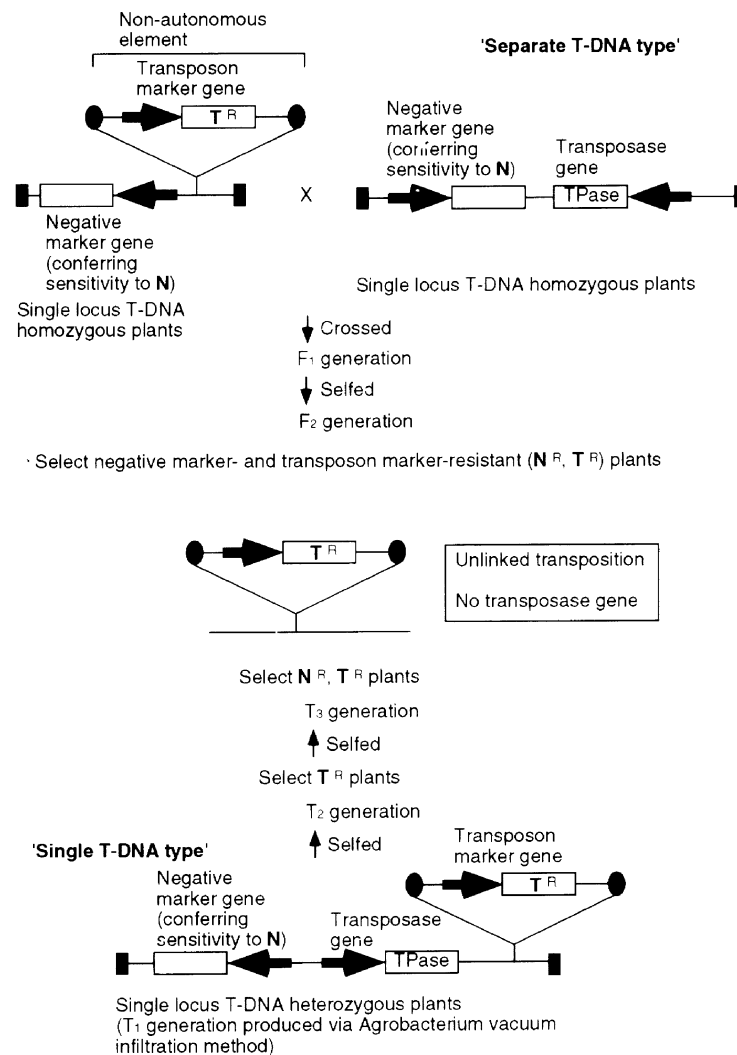


Figure 1.2 Construction and method for genome-wide transposition using two-component system

Both single (Tissier *et al.*, 1999) and separate (Sundaresan *et al.*, 1995) T-DNA types are successfully used. Symbols are the same as described in Figure 1.1. Transformation marker genes necessary for *Agrobacterium*-mediated transformation are omitted.

1.3. GENOME SATURATION WITH INSERTS

Bevan *et al.* (1998) estimated a gene density of one gene every 4.8-kb genomic DNA using 1.9 Mb of contiguous genomic sequence data. Based on this report, Krysan *et al.* (1999) estimated that median gene length was 2.1 kb, and ~280,000 inserts were necessary for 99% saturation, and 180,000 inserts for 95% saturation of the *Arabidopsis* genes.

To construct a saturated insertion library, two approaches must be considered. One is the use of plants with multiple copies of inserts per genome, and the other is the use of plants with a single copy or a few copies of inserts per genome. An advantage of the multicopy-insertional mutagenesis is that a relatively small number of plants is sufficient to saturate the whole *Arabidopsis* genome. For construction of this compact insertion library, Wisman *et al.* (1998) used an autonomous *En* element. This element was spontaneously multimerized in the *Arabidopsis* genome through 12 generations using a single seed descent procedure. Another group (Speulman *et al.*, 1999) adopted an *En/I* two-component system. Both systems contain multiple elements (~20 copies) per genome. However, multiple insertion systems using transposons make genetic and molecular analysis of mutants difficult. For example, somatic insertion may not be transmitted to the germ line, so the mutants may not be isolated. In other cases, successive transposition events often cause footprints when excised, and the footprints cause frame-shift mutations not related to the insertion.

On the other hand, single- or low-copy insertional mutagenesis makes the genetic and molecular analysis of mutants easier although a huge number of plants are necessary for saturation of the whole *Arabidopsis* genome. T-DNA insertion confers a few inserts per genome. Thus, T-DNA mutagenesis belongs to the low copy number insertional mutagenesis. To construct a transposon insertion library, which consists of a single or a few inserts per plant genome, two-component systems are used. In these systems, immobilized transposase genes are removed by segregation immediately after transposition has occurred during a single or two generations, and non-autonomous elements are stabilized (Fedoroff and Smith, 1993; Sundaresan *et al.*, 1995; Tissier *et al.*, 1999).

1.4. ENHANCER AND GENE TRAPPING

In recent years, enhancer trap and gene trap are playing important roles in the study of developmental biology. The enhancer trap element in a transposon or T-DNA carries a reporter gene fused to a minimal promoter, which can respond to endogenous *cis*-acting enhancer elements near the insertion site. The gene trap element carries a reporter gene fused to a splice acceptor sequence, and when inserted within an endogenous gene, it gives fused transcripts to the upstream exon in a genome. In both cases, the elements enable detection of genes by expression pattern even if the insertion does not confer mutant phenotype.

The single- or low-copy insertion library has an advantage for the screening by expression pattern using the enhancer or gene trap lines, because multiple insertion lines confer complex expression patterns derived from the multiple elements. In *Arabidopsis* mutagenesis, several groups adopted the enhancer or gene trap in their constructs, and constructed single- or low-copy insertion libraries (Fedoroff and Smith, 1993; Sundaresan *et al.*, 1995; Campisi *et al.*, 1999). From these libraries, many genes were isolated and analysed by using the enhancer and gene trap technologies (Smith and Fedoroff, 1995; Springer *et al.*, 1995; Tsugeki and Fedoroff, 1999; Ito *et al.*, 2000).

1.5. LOSS- AND GAIN-OF-FUNCTION MUTATIONS

When a T-DNA or transposon is inserted into an exon region of a gene, the gene will be disrupted. When inserted into an intron, the gene will be often disrupted because the insert is large enough compared with introns in *Arabidopsis*. These insertion mutants generally confer loss-of-function alleles. However, genome sequencing of various organisms revealed that there are many redundant genes with the same or similar functions. Indeed, it is estimated that over two thirds of genes in flies (*Drosophila melanogaster*), worms (*Caenorhabditis elegans*), and yeast

(*Saccharomyces cerevisiae*) have no obvious loss-of-function mutant phenotypes (Burns *et al.*, 1994; Miklos and Rubin, 1996). One approach to dissect redundant gene function is to disrupt all the redundant genes. For example, individual loss-of-function mutants of functionally redundant *ETR1*, *ETR2*, *EIN4* and *ERS2* genes, which are considered as ethylene receptor genes, showed no obvious ethylene response defects. However, a quadruple mutant of these single mutations exhibited a strong and constitutive ethylene response (Hua and Meyerowitz, 1998). Another example is functionally redundant *SEPALLATA 1*, *2* and *3* genes, which belong to MADS-box-type transcription factors. A triple mutant of these insertion mutations exhibited a mutant phenotype, in which all flower organs developed as sepals (Pelaz *et al.*, 2000). Another way to link redundant genes and functions is to analyze over- or misexpression phenotypes. In *Arabidopsis*, an overexpression mutant library known as activation tagging library has been constructed (Wilson *et al.*, 1996; Weigel *et al.*, 2000). In this system, the T-DNA or transposon contains tandem copies of a strong and constitutive CaMV 35S enhancer. When inserted near a gene, the enhancer can cause overexpression of the gene, and the dominant phenotype will appear. This approach has been used to identify genes involved in dominant *Arabidopsis* mutations such as *CYTOKININ-INDEPENDENT1*, *TINY*, and *SHORT-INTERNODES* (Kakimoto, 1996; Wilson *et al.*, 1996; Fridborg *et al.*, 1999). In *Arabidopsis*, misexpression screening has not been reported yet, but in *Drosophila*, a misexpression screening system using DNA-binding Gal4 transactivator has been developed (Rorth, 1996). In this system, tandem copies of Gal4 binding sites are contained within a transposable *P* element. When inserted near a gene, this element itself confers no effect. However, by crossing with 'a pattern line' in which the Gal4 gene controlled by a tissue-specific promoter was transformed, the gene near the *P* element becomes ectopically and tissue-specifically active. This screen has an advantage that the modular design of Gal4 system makes the screen flexible and conditional. The genes can be induced in any spatial or temporal pattern, and dominant lethal and sterile mutations can be recovered.

1.6. DETERMINATION OF THE INSERTION SITES

After construction of the random insertion library, the insertion sites have to be determined. Two major strategies are available for this purpose. One is a PCR-based screening for isolation of genes of interest, the other is precise mapping of the individual inserts.

1.6.1. PCR-BASED SCREENING

This strategy is successfully used for screening large populations of transposon or T-DNA insertion lines (McKinney *et al.*, 1995; Krysan *et al.*, 1999; Speulman *et al.*, 1999; Tissier *et al.*, 1999). An insertion within any given gene can be easily identified. PCR is performed using a primer set of gene-specific and insert-specific primers, and genomic DNA pooling from hundreds to thousands of plants. PCR product can be detected only when an insert exists within or close to the genes of interest.

1.6.2. MAPPING OF THE INDIVIDUAL INSERTS

By the end of the year 2000, the whole genomic DNA sequence of *Arabidopsis* will be determined. Determination of the genomic sequences that flank the inserts enables precise mapping of the inserts upon reference to published genomic DNA sequences. This strategy has become practical due to the development of effective methods for amplification of the flanking sequence such as TAIL-PCR (Liu and Whittier, 1995), inverse PCR (Ochman *et al.*, 1993), ligation-mediated PCR (Yephremov and Saedler, 2000), and plasmid rescue (Hayashi *et al.*, 1992). In *Arabidopsis*, it is estimated that the gene density is one gene every 4.8 kb genomic DNA, and the median gene length is 2.1 kb (Bevan *et al.*, 1998; Krysan *et al.*, 1999). This means that one out of two inserts on the average causes gene disruption. Because of the nature of high gene density in the *Arabidopsis* genome, an insertion site mapping project is now in progress (Ito *et al.*, 1999; Parinov *et al.*, 1999; Speulman *et al.*, 1999; Tissier *et al.*, 1999). For this purpose, transposable elements have an advantage, because those can be remobilized germinally and mapped insertion lines

will become useful resources for the targeted mutagenesis of closely linked genes. If one finds an insertion near a gene of interest, the gene disruptant may be obtained by a secondary transposition procedure. At saturation, screening for insertion mutants would be performed *in silico*.

1.7. INSERTIONAL MUTAGENESIS OF LARGER GENOME

Arabidopsis has a compact genome and high gene density. Because of this nature, construction of an insertion library which consists of a single or a few insert per genome is reasonable. However, in the case of other plant species with a larger genome size and lower gene density, it is difficult to prepare such insertion libraries. For example, in maize, at least 50% of the genome consists of nested retrotransposons, and these complex repeats in the intergenic regions range in size from 20 to 200 kb (SanMiguel *et al.*, 1996). This means that it is almost impossible to saturate maize genes by using single-copy insert plants. Alternatively, a compact insertion library which consists of multicopy inserts per genome may be practical. In petunia, maize, and snapdragon, endogenous multicopy transposons have been widely used for genetics. For reverse-genetical purposes in petunia, a PCR screening method to obtain insertion mutants has been developed using an endogenous high-copy-number *dTph1* transposon (Koes *et al.*, 1995). In maize, functional genomics projects are now cataloging transposon insertion sites to define all maize genes. The *Ac/Ds* and Robertson's mutator (*MuDR/Mu*) elements have been used in recent mutagenesis in maize. *Ac/Ds* is an endogenous low-copy-number transposon and an inefficient global mutagen, but efficient for local mutagenesis. On the other hand, *MuDR/Mu* is an efficient global mutagen because of its high copy number, but few germinal revertants are obtained (Walbot, 2000). In rice, there is no well-characterized endogenous transposon. Instead, for a reverse genetical approach, tissue culture-induced activation of an endogenous retrotransposon, *Tos17*, has been used to develop the insertional mutagenesis system (Hirochika, 1997). Recently, T-DNA insertion library of rice which contained ~25,000 inserts has been reported (Jeon *et al.*, 2000).

For insertional mutagenesis of the species with a low gene density, gene identification is also important. Since transposons prefer local transposition, it is important to identify genes near the transposon donor locus when transposons are used as insertional mutagens. For this purpose, we have developed a cDNA scanning method to concentrate cDNAs in a specific genome region from a total cDNA library (Hayashida *et al.*, 1995). In the cDNA scanning method, a mapped genomic DNA fragment covalently bound to latex beads is hybridized to PCR products of DNA inserts derived from a cDNA library. Washing non-hybridized fragments away leaves only cDNAs derived from the genomic DNA. We have applied this technique to the well-characterized *Arabidopsis* genome, and found it to be reliable (Seki *et al.*, 1997; Motohashi *et al.*, 1999). By combining this technique and *Ac/Ds* local transposition system, we have isolated 5 *Ds* insertion mutants of *Arabidopsis* near the transposon donor locus by PCR-based screening of 800 lines (Ito *et al.*, 1999; Seki *et al.*, 1999). This technique will be of use for species of lower gene density to isolate insertion mutants of genes near transposon donor locus.

1.8. CONCLUSIONS

In recent years, *Arabidopsis* insertion libraries have been extensively constructed in many laboratories and have almost reached saturation, although not all libraries are available for public use at present. As the next stage of *Arabidopsis* functional genomics, insertion mutants of already identified and predicted genes are being now extensively searched by using PCR-based screening. For example, insertions into 36 distinct R2R3 class *MYB* transcription family genes, which constitute a large family of more than 92 members, have already been isolated (Meissner *et al.*, 1999). Genome-wide expression analysis of the total *Arabidopsis* genes has also just started. Transcriptome (DNA microarray or DNA chip) and proteome techniques will be useful to list up affected downstream genes at the mRNA and protein levels, respectively, in the insertion mutants, although at present it is difficult to analyse genes expressed transiently or in a restricted tissue. These analyses will lead to our

understanding of the function and position of the individual genes in genetic networks.

ACKNOWLEDGMENT

We thank Reiko Motohashi for her helpful comments.

REFERENCES

- Aoyama, T., Dong, C.-H., Wu, Y., Carabelli, M., Sessa, G., Ruberti, I., Morelli, G. and Chua, N.-H. (1995) Ectopic expression of the *Arabidopsis* transcriptional activator *Athb-1* alters leaf cell fate in tobacco, *Plant Cell* **7**, 1773-1785.
- Balcells, L., Sundberg, E. and Coupland, G. (1994) A heat-shock promoter fusion to the *Ac* transposase gene drives inducible transposition of a *Ds* element during *Arabidopsis* embryo development, *Plant J.* **5**, 755-764.
- Bechtold, N., Ellis, J. and Pelletier, G. (1993) *In planta Agrobacterium* mediated gene transfer by infiltration of adult *Arabidopsis thaliana* plants, *C. R. Acad. Sci. Paris* **316**, 1194-1199.
- Bevan, M., *et al.* (1998) Analysis of 1.9 Mb of contiguous sequence from chromosome 4 of *Arabidopsis thaliana*, *Nature* **391**, 485-488.
- Burns, N., Grimwade, B., Ross-Macdonald, P.B., Choi, E.-Y., Finberg, K., Roeder, G.S. and Snyder, M. (1994) Large-scale analysis of gene expression, protein localization, and gene disruption in *Saccharomyces cerevisiae*, *Genes Dev.* **8**, 1087-1105.
- Campisi, L., Yang, Y., Yi, Y., Heilig, E., Herman, B., Cassista, A.J., Allen, D.W., Xiang, H. and Jack, T. (1999) Generation of enhancer trap lines in *Arabidopsis* and characterization of expression patterns in the inflorescence, *Plant J.* **17**, 699-707.
- Carpenter, R. and Coen, E.S. (1991) Floral homeotic mutations produced by transposon mutagenesis in *Antirrhinum majus*, *Genes Dev.* **4**, 1483-1493.
- Castle, L.A., Erramplli, D., Atherten, T.L., Franzmann, L.H., Yoon, E.S. and Meinke, D.W. (1993) Genetic and molecular characterization of embryonic mutations identified following seed transformation in *Arabidopsis*, *Mol. Gen. Genet.* **241**, 504-514.
- Clough, S.J. and Bent, A.F. (1998) Floral dip: a simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*, *Plant J.* **16**, 735-743.
- Fedoroff, N.V. and Smith, D.L. (1993) A versatile system for detecting transposition in *Arabidopsis*, *Plant J.* **3**, 273-289.
- Feldmann, K.A. (1991) T-DNA insertion mutagenesis in *Arabidopsis*: Mutational spectrum, *Plant J.* **1**, 71-82.
- Feldmann, K.A. and Marks, M.D. (1987) *Agrobacterium*-mediated transformation of germinating seeds of *Arabidopsis thaliana*: a non-tissue culture approach, *Mol. Gen.*

Genet. **208**, 1-9.

- Fridborg, I., Kuusk, S., Moritz, T. and Sundberg, E. (1999) The Arabidopsis dwarf mutant *shi* exhibits reduced gibberellin responses conferred by overexpression of a new putative zinc finger protein, *Plant Cell* **11**, 1019-1031.
- Gheysen, G., Herman, L., Breyne, P., Gielen, J., Van Montagu, M. and Depicker, A. (1990) Cloning and sequence analysis of truncated T-DNA inserts from *Nicotiana tabacum*, *Gene* **94**, 155-163.
- Grevelding, C., Becker, D., Kunze, R., Von Menges, A., Fantes, V., Schell, J. and Masterson, R. (1992) High rates of Ac/Ds germinal transposition in Arabidopsis suitable for gene isolation by insertional mutagenesis, *Proc. Natl. Acad. Sci. USA* **89**, 6085-6089.
- Hayashi, H., Czaja, I., Lubenow, H., Schell, J. and Walden, R. (1992) Activation of a plant gene by T-DNA tagging: auxin-independent growth *in vitro*, *Science* **258**, 1350-1353.
- Hayashida, N., Sumi, Y., Wada, T., Handa, H. and Shinozaki, K. (1995) Construction of a cDNA library for a specific region of a chromosome using a novel cDNA selection method utilizing latex particles, *Gene* **165**, 155-161.
- Hirochika, H. (1997) Retrotransposons of rice: their regulation and use for genome analysis, *Plant Mol. Biol.* **35**, 231-240.
- Hofte, H. *et al.* (1993) An inventory of 1152 expressed sequence tags obtained by partial sequencing of cDNAs from Arabidopsis thaliana, *Plant J.* **4**, 1051-1061.
- Hua, J. and Meyerowitz, E.M. (1998) Ethylene responses are negatively regulated by a receptor gene family in *Arabidopsis thaliana*, *Cell* **94**, 261-271.
- Ito, T., Kim, G.-T. and Shinozaki, K. (2000) Disruption of an *Arabidopsis* cytoplasmic ribosomal protein S13-homologous gene by transposon-mediated mutagenesis causes aberrant growth and development, *Plant J.* **22**, 257-264.
- Ito, T., Seki, M., Hayashida, N., Shibata, D. and Shinozaki, K. (1999) Regional insertional mutagenesis of genes on *Arabidopsis thaliana* chromosome V using the *Ac/Ds* transposon in combination with a cDNA scanning method, *Plant J.* **17**, 433-444.
- Jeon, J.-S. *et al.* (2000) T-DNA insertional mutagenesis for functional genomics in rice, *Plant J.* **22**, 561-570.
- Kakimoto, T. (1996) CKI1, a histidine kinase homolog implicated in cytokinin signal transduction, *Science* **274**, 982-985.
- Kempin, S.A., Liljegren, S.J., Block, L.M., Rounsley, S.D., Yanofsky, M.F. and Lam, E. (1997) Targeted disruption in Arabidopsis, *Nature* **389**, 802-803.
- Konieczny, A., Voytas, D.F., Cummings, M.P. and Ausubel, F.M. (1991) A superfamily of *Arabidopsis thaliana* retrotransposons, *Genetics* **127**, 801-809.
- Krysan, P.J., Young, J.C. and Sussman, M.R. (1999) T-DNA as an insertional mutagen in Arabidopsis, *Plant Cell* **11**, 2283-2290.
- Lin, X. *et al.* (1999) Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana, *Nature* **402**, 761-768.

- Liu, Y.G. and Whittier, R.F. (1995) Thermal asymmetric interlaced PCR: automatable amplification and sequencing of insert end fragments from P1 and YAC clones for chromosome walking, *Genomics* **25**, 674-681.
- Machida, C., Onouchi, H., Koizumi, J., Hamada, S., Semiarti, E., Torikai, S. and Machida, Y. (1997) Characterization of the transposon pattern of the *Ac* element in *Arabidopsis thaliana* using endonuclease I-SceI, *Proc. Natl. Acad. Sci. USA* **94**, 8675-8680.
- Mayer, K. *et al.* (1999) Sequence and analysis of chromosome 4 of the plant *Arabidopsis thaliana*, *Nature* **402**, 769-777.
- McKinney, E.C., Ali, N., Traut, A., Feldmann, K.A., Belostotsky, D.A., McDowell, J.M. and Meagher, R.B. (1995) Sequence-based identification of T-DNA insertion mutations in *Arabidopsis*: actin mutants *act2-1* and *act4-1*, *Plant J.* **8**, 613-622.
- Meissner, R.C. *et al.* (1999) Functional search in a large transcription factor gene family in *Arabidopsis*: assessing the potential of reverse genetics to identify insertional mutations in R2R3 MYB genes, *Plant Cell* **11**, 1827-1840.
- Miao, Z.-H. and Lam, E. (1995) Targeted disruption of the TGA3 locus in *Arabidopsis thaliana*, *Plant J.* **7**, 359-365.
- Miklos, G.L.G. and Rubin, G.M. (1996) The role of the genome project in determining gene function: insights from model organisms, *Cell* **86**, 521-529.
- Motohashi, R., Ito, T., Seki, M., Ichimura, K., Yamaguchi-Shinozaki, K. and Shinozaki, K. (1999) Evaluation of a cDNA scanning method concerning the fidelity and efficiency of cDNA selection using the YAC CIC3B1-S region of *Arabidopsis thaliana* chromosome 5, *DNA Res.* **6**, 247-253.
- Nacry, P., Camilleri, C., Courtial, B., Caboche, M. and Bouchez, D. (1998) Major chromosomal rearrangements induced by T-DNA transformation in *Arabidopsis*, *Genetics* **149**, 641-650.
- Newman, T., de Bruijn, F.J., Green, P., Keegstra, K., Kende, H., McIntosh, L., Ohlrogge, J., Raikhel, N., Somerville, S., Thomashow, M., Retzel, E. and Somerville, C. (1994) Genes galore: A summary of methods for accessing results from large-scale partial sequencing of anonymous *Arabidopsis* cDNA clones, *Plant Physiol.* **106**, 1241-1255.
- Ochman, H., Ayala, F.J. and Hartl, D.L. (1988) Use of polymerase chain reaction to amplify segments outside boundaries of known sequences, *Methods Enzymol.* **218**, 309-321.
- Oliver, S.G. (1996) From DNA sequence to biological function, *Nature* **379**, 597-600.
- Pelaz, S., Ditta, G.S., Baumann, E., Wisman, E. and Yanofsky, M.F. (2000) B and C floral organ identity functions require *SEPALLATA* MADS-box genes, *Nature* **405**, 200-203.
- Peleman, J., Cottyn, B., Van Camp, W., Van Montagu, M. and Inze, D. (1991) Transient occurrence of extrachromosomal DNA of an *Arabidopsis thaliana* transposon-like element, *Tat1*, *Proc. Natl. Acad. Sci. USA* **88**, 3618-3622.
- Pelissier, T., Tutois, S., Deragon, J.M., Tourmente, S., Genestier, S. and Picard, G.

- (1995) *Athla*, a new retroelement from *Arabidopsis thaliana*, *Plant Mol. Biol.* **29**, 441-452.
- Rorth, O. (1996) A modular misexpression screen in *Drosophila* detecting tissue-specific phenotypes, *Proc. Natl. Acad. Sci. USA* **93**, 12418-12422.
- SanMiguel, P., Tikhonov, A., Jin, Y.-K., Motchoulskaia, N., Zakharov, D., Melake-Berhan, A., Springer, P.S., Edwards, K.J., Lee, M., Avramova, Z. and Bennetzen, J.L. (1996) Nested retrotransposons in the intergenic regions of the maize genome, *Science* **274**, 765-768.
- Seki, M., Hayashida, N., Kato, N., Yohda, M. and Shinozaki, K. (1997) Rapid construction of a transcription map for a cosmid contig of *Arabidopsis thaliana* genome using a novel cDNA selection method, *Plant J.* **12**, 481-487.
- Seki, M., Ito, T., Shibata, D. and Shinozaki, K. (1999) Regional insertional mutagenesis of specific genes on the CIC5F11/CIC2B9 locus of *Arabidopsis thaliana* chromosome 5 using the *Ac/Ds* transposon in combination with the cDNA scanning method, *Plant Cell Physiol.* **40**, 624-639.
- Simon, R., Igeno, M.I. and Coupland, G. (1996) Activation of floral meristem identity genes in *Arabidopsis*, *Nature* **384**, 59-62.
- Smith, D.L. and Fedoroff, N.V. (1995) *LRP1*, a gene expressed in lateral and adventitious root primordia of *Arabidopsis*, *Plant Cell* **7**, 735-745.
- Speulman, E., Metz, P.L.J., van Arkel, G., Hekkert, B.L., Stiekema, W.J. and Pereira, A. (1999) A two-component *Enhancer-Inhibitor* transposon mutagenesis system for functional analysis of the *Arabidopsis* genome, *Plant Cell* **11**, 1853-1866.
- Springer, P.S., McCombie, W.R., Sundaresan, V. and Martienssen, R.A. (1995) Gene trap tagging of *PROLIFERA*, an essential *MCM2-3-5*-like gene in *Arabidopsis*, *Science* **268**, 877-880.
- Sundaresan, V., Springer, P., Volpe, T., Haward, S., Jones, J.D., Dean, C., Ma, H. and Martienssen, R. (1995) Patterns of gene action in plant development revealed by enhancer trap and gene trap transposable elements, *Genes Dev.* **9**, 1797-1810.
- Tissier, A.F., Marillonnet, S., Klimyuk, V., Patel, K., Torres, M.A., Murphy, G. and Jones, J.D.G. (1999) Multiple independent defective *Suppressor-mutator* transposon insertions in *Arabidopsis*: A tool for functional genomics, *Plant Cell* **11**, 1841-1852.
- Tsay, Y.-F., Frank, M.J., Page, T., Dean, C. and Crawford, N.M. (1993) Identification of a mobile endogenous transposon in *Arabidopsis thaliana*, *Science* **260**, 342-344.
- Tsugeki, R. and Fedoroff, N.V. (1999) Genetic ablation of root cap cells in *Arabidopsis*, *Proc. Natl. Acad. Sci. USA* **96**, 12941-12946.
- Walbot, V. (2000) Saturation mutagenesis using maize transposons, *Curr. Opin. Plant Biol.* **3**, 103-107.
- Weigel, D. et al. (2000) Activation tagging in *Arabidopsis*, *Plant Physiol.* **122**, 1003-1013.
- Wilson, K., Long, D., Swinburne, J. and Coupland, G. (1996) A Dissociation insertion causes a semidominant mutation that increases expression of *TINY*, an *Arabidopsis* gene related to *APETALA2*, *Plant Cell* **8**, 659-671.

- Wisman, E., Cardon, G.H., Fransz, P. and Saedler, H. (1998) The behaviour of the autonomous maize transposable element *En/Spm* in *Arabidopsis thaliana* allows efficient mutagenesis, *Plant Mol. Biol.* **37**, 989-999.
- Yephremov, A. and Saedler, H. (2000) Display and isolation of transposon-flanking sequences starting from genomic DNA or RNA, *Plant J.* **21**, 495-505.

ENGINEERING THE CHLOROPLAST GENOME TO CONFER STRESS TOLERANCE AND PRODUCTION OF PHARMACEUTICAL PROTEINS

Henry Daniell

*Department of Molecular Biology & Microbiology and Center for Discovery of Drugs and
Diagnostics, University of Central Florida, 12722 Research parkway, Orlando FL 32826-3227,
USA.*

The life-supporting metabolic process of photosynthesis is carried out in endosymbiotic organelles, the chloroplasts. Scientific endeavors over the last decade have enabled us to successfully engineer the chloroplast genome. This singular development has opened new avenues for agricultural biotechnology and has contributed immensely to our understanding of the basic genetic mechanisms operative within the organelle. The chloroplast genome has been engineered to express agronomically important traits such as disease-resistance, drought-tolerance, herbicide-resistance, insect resistance and production of antibodies, biopharmaceuticals and edible vaccines. The ability to hyper-express prokaryotic and eukaryotic proteins without the drawbacks of gene silencing or position effects, combined with the advantage of gene containment due to plastid maternal inheritance, make chloroplast transformation technology both useful and environmentally-friendly. The recent use of a plant-derived selectable marker for chloroplast transformation is an important development that should help allay public fears about GM foods. In addition, the extension of chloroplast transformation technology to edible plant species such as potato and tomato, coupled with the hyper-expression of antigens, opens up the possibility of oral delivery of vaccines and other biopharmaceuticals. This review focuses on some of these recent accomplishments and their impact.

Chloroplast genetic engineering is emerging as an alternative new technology to overcome some of the potential environmental concerns of nuclear genetic engineering (reviewed in ref 1). One commonly perceived environmental concern is the escape of foreign gene through pollen or seed dispersal from transgenic crop plants to their weedy relatives creating super weeds or causing genetic pollution among other crops (2). High rates of such gene flow from crops to wild relatives (as high as 38% in sunflower and 50%

for strawberries) are certainly a serious concern. Keeler et al. (3) have summarized valuable data on the weedy wild relatives of sixty important crop plants and potential hybridization between crops and wild relatives. Among sixty crops, only eleven do not have congeners (members of the same genus) and the rest of the crops have wild relatives somewhere in the world. In addition, genetic pollution among crops has resulted in several lawsuits and shrunk the European market for organic produce from Canada from 83 tons in 1994–1995 to 20 tons in 1997–1998 (4). Several major food corporations, including ADM have required segregation of native crops from those polluted with transgenes. Two legislations have been submitted in the U.S. to protect organic farmers whose crops inadvertently contain transgenes via pollen drift (5). Maternal inheritance of foreign genes through chloroplast genetic engineering is highly desirable in such instances where there is potential for out-cross among crops or between crops and weeds (6-8).

Yet another concern in the use of nuclear transgenic crops expressing the *Bacillus thuringiensis* (Bt) toxins is the sub-optimal production of toxins resulting in increased risk of pests developing Bt resistance. Plant-specific recommendations to reduce Bt resistance development include increasing Bt expression levels (high dose strategy), expressing multiple toxins (gene pyramiding), or expressing the protein only in tissues highly sensitive to damage (tissue specific expression). All three approaches are attainable through chloroplast genetic engineering. For example, hyper-expression of several thousand copies of a novel B.t. gene via chloroplast genetic engineering, resulted in 100% mortality of insects that are up to 40,000-fold resistant to other B.t. proteins (9). Another hotly debated environmental concern expressed recently is the toxicity of transgenic pollen to non-target insects, such as the Monarch butterflies (10,11). However, toxic insecticidal protein was not expressed in pollen of chloroplast transgenic plants, even though leaves expressed extraordinarily high levels of transgenic protein (12). Chloroplast gene expression also results in tissue specificity, occurring predominantly where functional plastids are present. This may be important in engineering insect resistant plants wherein most worms predominantly feed on leaves where plastids are abundantly present, thereby consuming the highest level of the insecticidal protein. If desired, regulatory signals specific for non-green plastids should be used to engineer insect resistance in roots, tubers or other organs.

Since the first recombinant product was approved for marketing in 1982, the biopharmaceutical industry has undergone a dramatic transition. Currently genes are routinely cloned and their derivative proteins expressed in bacteria, insects and mammalian cells for commercial purposes. The FDA approved 26 new biopharmaceuticals in the five-year period of 1995 to 1999. Current trends imply that the demand and cost of biopharmaceuticals is expected to escalate exponentially in the foreseeable future. Therefore it is important to explore cost-effective methods of biopharmaceutical production. Plants have emerged as a promising alternative for the production of

recombinant proteins due to unique advantages they offer. In contrast to the industrial systems employing fermenters or bioreactors, plant based systems are relatively inexpensive. Establishment of large scale harvesting or downstream processing procedures is not new for these systems because this technology has been an integral part of agriculture for many years now. Expression of a particular protein in plants can also eliminate the need for its purification, if there is a possibility of consuming it directly. This also minimizes the health risks arising from contamination with potential human pathogens or toxins. Plants also provide the choice of compartmentalized production of recombinant proteins either via targeting or direct expression in the sub-cellular organelles where foreign proteins are protected from proteolytic degradation.

A remarkable feature of chloroplast genetic engineering is the observation of exceptionally large accumulation of foreign proteins in transgenic plants, including more than 50% of CRY protein in total soluble protein, even in bleached old leaves (12, 13). Stable expression of a pharmaceutical protein in chloroplasts was first reported for GVGVP, a protein based polymer with varied medical applications (such as the prevention of post-surgical adhesions and scars, wound coverings, artificial pericardia, tissue reconstruction and programmed drug delivery, 14). Subsequently, expression of the human somatotropin via the tobacco chloroplast genome (15) to high levels (7% of total soluble protein) was observed. Since then, the transgenic chloroplast genome has been successfully engineered to produce complex bacterial proteins that require oligomeric assembly and establishment of disulfide bonds (cholera toxin β subunit), human serum albumin that requires seventeen disulfide bonds and monoclonals that require assembly of multimeric proteins. It is well known that the level of foreign gene expression is not adequate for commercial feasibility of several pharmaceutical proteins when expressed via the nuclear genome; levels of expression of pharmaceutical proteins vary over three orders of magnitude, 0.001 to 1% of total soluble protein (16). Therefore, it is wise to exploit this major advantage by engineering foreign genes via the chloroplast genome instead of the nuclear genome. Hyperexpression at the site of infection is also highly desirable to control invasion of pathogens in transgenic plants. Because of the concentration dependent action of the anti-microbial peptide in controlling bacterial infection, we expressed it via the chloroplast genome to accomplish high mortality of bacteria and fungi at the point of infection (17). Chloroplast transformation utilizes two flanking sequences that, through homologous recombination, insert foreign DNA into the spacer region between the functional genes of the chloroplast genome, thus targeting the foreign genes to a precise location. This eliminates the "position effect" and gene silencing observed in nuclear transgenic plants (18,19).

In plant and animal cells, nuclear mRNAs are translated monocistronically. This poses a serious problem when engineering multiple genes in plants (1). Therefore, in order to express the polyhydroxybutyrate polymer or Guy's 13 antibody, single genes were first introduced into

individual transgenic plants, then these plants were back-crossed to reconstitute the entire pathway or the complete protein (20,21). Similarly, in a seven yearlong effort, Ye et al (22) recently introduced a set of three genes for a short biosynthetic pathway that resulted in β -carotene expression in rice. In contrast, most chloroplast genes of higher plants are contrascribed (1). Multiple steps of chloroplast mRNA processing are involved in the formation of mature mRNAs. Expression of polycistrons via the chloroplast genome provides a unique opportunity to express entire pathways in a single transformation event. An example of expression of a bacterial operon in plants via the chloroplast genome will be discussed in this review (12).

There have been several efforts to generate various stress resistant transgenic plants by introducing gene(s) responsible for trehalose biosynthesis, regulation or degradation (23-25). When trehalose accumulation was increased in transgenic tobacco plants by over-expression of the yeast *TPS1*, trehalose accumulation resulted in the loss of apical dominance, stunted growth, lancet shaped leaves and some sterility. Altered phenotype was always correlated with drought tolerance; plants showing severe morphological alterations had the highest tolerance under stress conditions. Several toxic compounds expressed in transgenic plants have been compartmentalized in chloroplasts even through no targeting sequence was provided (26,27), indicating that this organelle could be used as a repository like the vacuole. Also, osmoprotectants are known to accumulate inside chloroplasts under stress conditions (29). Inhibition of trehalase activity is known to enhance trehalose accumulation in plants (24). Therefore, trehalose accumulation in chloroplasts may be protected from trehalase activity in the cytosol, if trehalase was absent in the chloroplast. In order to minimize the pleiotropic effects observed in the nuclear transgenic plants accumulating trehalose, a recent study attempted to compartmentalize trehalose accumulation within chloroplasts. An example of trehalose accumulation inside chloroplasts and resultant drought tolerant phenotypes (28) is discussed in this review.

Most transformation techniques co-introduce a gene that confers antibiotic resistance, along with the gene of interest to impart a desired trait. Regenerating transformed cells in antibiotic containing growth media permits selection of only those cells that have incorporated the foreign genes. Once transgenic plants are regenerated, antibiotic resistance genes serve no useful purpose but they continue to produce their gene products. One among the primary concerns of genetically modified (GM) crops is the presence of clinically important antibiotic resistance gene products in transgenic plants that could inactivate oral doses of the antibiotic (reviewed in 30; 31). Alternatively, the antibiotic resistant genes could be transferred to pathogenic microbes in the gastrointestinal tract or soil rendering them resistant to treatment with such antibiotics. Antibiotic resistant bacteria are one of the major challenges of modern medicine. In Germany, GM crops containing antibiotic resistant genes have been banned from release (32). However,

several approaches are currently available to eliminate antibiotic resistance genes from nuclear transgenic crops (30). An example of marker-free chloroplast genetic engineering will be discussed in this review.

Engineering Herbicide Resistance via the Chloroplast Genome

Selective herbicides are routinely applied to control weeds that would otherwise compete for available nutrients, space and light, thereby reducing crop yield. For example, glyphosate is a potent, broad spectrum herbicide which is highly effective against a majority of grasses and broad leaf weeds. Glyphosate works by competitive inhibition of the enzyme 5-enol-pyruvyl shikimate-3-phosphate synthase (EPSPS) of the aromatic amino acid biosynthetic pathway. Synthesis of EPSP from shikimate 3-phosphate and inorganic phosphate is catalyzed by EPSPS. This particular reaction occurs only in plants and microorganisms which explains the non-toxicity of glyphosate to other living forms. Use of glyphosate is environmentally safe as it is inactivated rapidly in soil, has minimum soil mobility, and degrades to natural products, with little toxicity to non-plant life forms. However, glyphosate lacks selectivity and does not distinguish crops from weeds, thereby restricting its use. EPSPS based glyphosate resistance has been genetically engineered by the overproduction of the wild type EPSPS (33) or by the expression of a mutant gene (*aroA*) encoding glyphosate resistant EPSPS (34). In all of the aforementioned examples, without exception, herbicide resistant genes have been introduced into the nuclear genome.

One common environmental concern is the escape of a foreign gene through pollen or seed dispersal, thereby creating super weeds or causing genetic pollution among other crops. Escape of herbicide resistance genes to wild relatives occurs predominantly via dispersal of viable pollen. Keeler et al. (3) focus on the role of gene flow to weedy wild relatives as a potential problem because in their opinion "this is a far greater concern than any other mode of escape of transgenes." These authors further point out that "transgenes can only reach weed populations if carried to weeds on viable pollen; if the crop produces no pollen or viable pollen, there will be no gene flow." The potential for gene flow via pollen depends on several factors including the amount of pollen produced, longevity of pollen, dispersal of pollen (via wind, animal), plant/weed density, dormancy/rehydration of pollen, survival of pollen from toxic substances secreted by pollinators and distance between crops and weeds. Keeler et al. (3) point out that it is impractical to prevent out-cross between weeds and wind pollinated crops because of the large pollen clouds produced and distance traveled by viable pollen.

However, it is possible, under exceptional circumstances, for the herbicide resistance crop to be fertilized by pollen from wild relatives and serve as a female parent for a hybrid seed. If this happens the hybrid seed

may germinate and establish a resistant population. However, for this to happen, the herbicide resistant crop that served as the female parent must escape harvesting and the hybrid seeds must survive to germinate, grow and reproduce. Alternatively, dispersal of seeds from transgenic plants may occur among weedy relatives, during harvest, transportation, planting and harvest. This can give rise to mixed populations. Introgressive hybridization could result in super weeds. This again would depend on the persistence of the crop among weeds and probability of forming mixed strands.

Genetic containment methods include apomixis, incompatible genomes, transgenic mitigation, control of seed dormancy, seed ripening or shattering, suicide genes, infertility barriers, male sterility and maternal inheritance (35). The latter two have been experimentally tested. Anther, the male reproductive organ, is composed of several cell and tissue types and contains anther specific mRNA's (36). Anther produces pollen grains that contain sperm cells. A specialized anther tissue called the tapetum plays an important role in the formation of pollen. The tapetum generally surrounds the pollen sac in early development and is not present as an organized tissue in the mature anther. The tapetum synthesizes a number of proteins that aid in pollen development or become components of pollen. Many male sterility mutations interfere with the tapetal cell differentiation and/or function, indicating that this tissue is essential for the production of functional pollen. Mariani et al. (36) have shown that the 5'-region of a tobacco tapetum-specific gene (TA29) can activate the expression of β -glucuronidase and ribonuclease genes (RNase T1 and barnase) within the tapetal cells of transgenic tobacco and oil seed rape plants. Expression of RNase genes selectively destroyed the tapetum during anther development, prevented pollen formation and produced male sterile plants. This approach could be used to contain out-cross of transgene with other crops or weeds. However, male sterility is possible only in crops where the product is not a seed or fruit requiring fertilization (like lettuce, carrot or cabbage).

Scott and Wilkinson (7) have recently analyzed several factors that would influence the transgene movement of chloroplast genes from crops to wild relatives under natural conditions. They studied the mode of inheritance of plastids, incidence of sympatry to quantify opportunities for forming mixed populations and persistence of crops outside agriculture limits for introgression. They studied plastid inheritance in natural hybrids collected from two wild populations growing next to oilseed rape along 34 km of the Thames River and assessed the persistence of 18 feral oil seed rape populations over a period of three years. These studies concluded that there would be no pollen-mediated transgene movement from oilseed rape. A low incidence of sympatry (0.6-0.7%) between the crop and weed species occurred; however, mixed strands showed a strong tendency towards rapid decline in plant number, seed return and ultimately extinction within three years. Thus, they concluded that gene flow will be rare if plants are genetically engineered via the chloroplast genome.

The prevalent pattern of plastid inheritance found in the majority of angiosperms is uniparental maternal and chloroplast genomes are maternally inherited for most of the crops. However, there are always exceptions to most observations and maternal inheritance of chloroplast genomes is certainly not without exception. It is known that in pines (gymnosperms) and a few flowering plants (like alfalfa) plastids are transmitted in a biparental mode. Paternal transmission of pollen in tobacco has been reported, but with provisos. In transmission of paternal chloroplasts in tobacco, authors mention that there is occasional (0.07-2.5%) paternal transmission in a species typically exhibiting strict maternal inheritance (see ref 37).

Maternal inheritance of a herbicide resistance gene and prevention of escape via pollen has been successfully demonstrated recently (6). Engineering foreign genes through chloroplast genomes (which are maternally inherited in most of the crops) is a practical solution to this problem. In addition, the target enzymes or proteins for most herbicides (of the amino acid / fatty acid biosynthetic pathways or photosynthesis) are compartmentalized within the chloroplast. Because the transcriptional and translational machinery of the chloroplast is prokaryotic in nature, herbicide resistant genes of bacterial origin can be expressed at extraordinarily high levels in chloroplasts. An example of genetic engineering herbicide resistance via chloroplast genome to overcome out-cross and gene pollution problems is discussed below.

The chloroplast vector pZS-RD-EPSPS contained the 16S rRNA promoter (Prn) driving the *aadA* (aminoglycoside adenylyl transferase) and EPSPS genes with the *psbA* 3' region (the terminator from a gene coding for photosystem II reaction center components) from the tobacco chloroplast genome. This construct integrated the EPSPS and *aadA* genes into the spacer region between the *rbcL* (the gene for the large subunit of RuBisCO) and *orf512* genes (code for the *accD* gene) of the tobacco chloroplast genome. This vector is useful to integrate foreign genes specifically into the tobacco chloroplast genome; this gene order is not conserved among other plant chloroplast genomes (38). On the other hand, the universal chloroplast expression and integration vector pSBL-RD-EPSPS can be used to transform chloroplast genomes of several other plant species because the flanking sequences are highly conserved among higher plants; the universal vector uses *trnA* and *trnI* genes (chloroplast transfer RNAs coding for Alanine and Isoleucine) from the inverted repeat region of the tobacco chloroplast genome as flanking sequences for homologous recombination.

Transgenic plants were obtained within 3-5 months after bombardment as described by Daniell (39,40). The integration of the *aroA* gene into the chloroplast was confirmed by PCR and Southern analyses. In addition, the high level of resistance to glyphosate observed, was confirmed by determination of the copy number of the foreign gene, in the transgenic plants. The copy number of the integrated gene was determined by establishing homoplasmy for the transgenic chloroplast genome. Tobacco

Chloroplasts contain 5000~10,000 copies of their genome per cell. If only a fraction of the genomes were actually transformed, the copy number, by default, must be less than 10,000. By establishing that in the transgenics the EPSPS transformed genome was the only one present, one could establish that the copy number is 5000~10,000 per cell. This proved that only the transgenic chloroplast genome was present in the cell and there was no native, untransformed, chloroplast genome, without the EPSPS gene present. This established the homoplasmic nature of transformants, simultaneously providing an estimate of about 10,000 copies of the foreign EPSPS gene per cell. This explained the high levels of tolerance to glyphosate observed in transgenic plants.

Seeds collected from transgenic plants after the first self-cross were germinated in the presence of spectinomycin. All of the seeds germinated, remained green and grew normally. The 100% resistance to spectinomycin in all of the clones examined showed maternal inheritance of the introduced genes. A heteroplasmic condition would have given rise to variegated progeny on spectinomycin; lack of such variegated progeny also confirms homoplasmy as confirmed by Southern blot analysis. All of the untransformed seedlings were bleached and did not grow in the presence of spectinomycin. Lack of variation in chlorophyll pigmentation among the progeny also underscores the absence of position effect, an artifact of nuclear transformation.

Eighteen week old control and transgenic plants were sprayed with equal volumes of different concentrations (0.5 to 5 mM) of glyphosate. Untransformed control tobacco plants were extremely sensitive to glyphosate; they died within seven days even at 0.5 mM glyphosate. On the other hand, the chloroplast transgenic plants survived concentrations as high as 5mM glyphosate. These results are intriguing, considering the fact that the EPSPS gene from petunia used in these chloroplast vectors, has a low level of tolerance to glyphosate. Sensitivity to glyphosate by EPSPS should have been compensated by overproduction of the enzyme by thousands of copies of the EPSPS gene, present in each cell of the transgenic plants. Also, this is the first report of expressing a eukaryotic nuclear gene within the prokaryotic chloroplast compartment. It is well known that the codon preference is significantly different between the prokaryotic chloroplast compartment and the eukaryotic nuclear compartment. Ideally, a mutant *aroA* gene from a prokaryotic system (which does not bind glyphosate) should be expressed in the chloroplast compartment; such genes are now available and exhibit a thousand fold higher level of resistance to glyphosate than the petunia gene used in this investigation. In light of these observations, it is possible that integration of prokaryotic herbicide resistance genes into the chloroplast genome could result in incredibly high levels of resistance to herbicides while still maintaining the efficacy of biological containment.

Engineering Insect Resistance via the Chloroplast Genome

The use of commercial, nuclear transgenic crops expressing *Bacillus thuringiensis* (Bt) toxins has escalated in recent years due to their advantages over traditional chemical insecticides. However, in crops with several target pests, each with varying degrees of susceptibility to Bt (e.g. cotton), there is concern regarding the sub-optimal production of toxin, resulting in reduced efficacy and increased risk of Bt resistance. Additionally, reliance on a single (or similar) B.t. protein(s) for insect control increases the likelihood of B.t.-resistance development (41). Most current commercial transgenic plants that target lepidopteran pests contain either Cry1Ab (corn) or Cry1Ac (cotton) (42,43). Bt corn is targeted primarily against European corn borer although other pests such as the corn earworm or cotton bollworm may be affected. B.t. cotton is targeted primarily against the tobacco budworm; however, other pests such as armyworms and cotton bollworm are economically damaging, but have only limited susceptibility to Cry1Ac.

Use of single Bt proteins to control insects such as tobacco budworm and cotton bollworm could lead to relatively rapid Bt resistance development (44,45). Additionally, because Cry1Ab and Cry1Ac share over 90% protein homology, resistance to one Cry1A protein would most likely impart resistance to another Cry1A protein as has been observed in tobacco budworm (45,46). Nowhere is this more of a concern than with cotton bollworm/ corn ear worm which usually feeds on corn in the spring and early summer, then migrates over to cotton to complete several more generations (44). Clearly, different Bt proteins are needed in order to decrease the development of resistance. The primary strategy currently used to delay development of insect resistance to *Bt* plants is to provide refuges of host plants that do not produce B.t. toxins. However, a recent study (47) of a resistant strain of pink bollworm larvae on B.t. cotton shows developmental asynchrony - this favors assortative mating among resistant moths emerging from B.t. plants, and generates a disproportionately high number of homozygous resistant insects, accelerating the evolution of B.t. resistance.

Another environmental concern expressed recently is the toxicity of transgenic pollen to non-target insects, including Monarch butterflies (10) although this study has been criticized as being premature and incomplete (11). Because there is no chloroplast DNA in the pollen of most crops, toxic insecticidal proteins should not be expressed in pollen of chloroplast transgenic plants (12). Evolving levels of B.t.-resistance in insects should be dramatically reduced through the genetic engineering of the chloroplast genome in transgenic plants. Therefore, one such example of chloroplast genetic engineering is discussed below.

The tobacco chloroplast expression vector described above was also used to introduce a novel B.t. coding sequence into the chloroplast genome. This class of Bt proteins, Cry2A, is toxic to many caterpillars, such as the European corn borer and tobacco budworm, is quite different in structure/function from the

Cry1A proteins (resulting in less cross resistance). Cry2A proteins are about half the size of Cry1A proteins, and therefore should be expressed at higher levels. Tobacco leaves were bombarded with DNA-coated tungsten particles as described elsewhere (39,40). The positive clones were analyzed by PCR and Southern hybridization to confirm the site-specific integration of *cry2Aa2*, and to establish copy number as explained before. Insect bioassays resulted in the following observations (9). There was 100% mortality of tobacco budworm feeding on transgenic leaves and the leaf pieces were essentially intact, while the control leaf pieces were completely devoured. Similar results were obtained with Cry1Ac and Cry2A resistant insects. Bioassays also were conducted using insects that were reared on control leaves or artificial diet for 5 days (ca. 2nd-3rd instar), and then moved to transgenic leaves. Even these older larvae that are more tolerant than neonates, showed 100% mortality. When transgenic leaves were fed to cotton bollworm and beet armyworm 100% mortality was observed, whereas there was no mortality observed in the control, and the entire leaf piece was devoured (9).

With the successful introduction of *cry2Aa2* into the chloroplast genome, the high-dose strategy should be attainable in other crops. This study shows 100% mortality of both Bt susceptible and Cry1Ac-resistant and Cry2Aa2-resistant tobacco budworm. This is the first report where neonate insects, highly resistant to Bt, were killed using Bt transgenic leaf material even though tobacco budworm is less sensitive to Cry2Aa2 than Cry1Ac. These results are promising when related to reports showing marginal to high levels of cross-resistance to Cry2Aa2 (45,46). This study also shows 100% mortality of cotton bollworm that contrasts with Bt cotton (Cry1Ac) efficacy against cotton bollworm. The inefficient control of cotton bollworm might also result in faster development of Bt resistance because a moderate level of suppression (25-50% mortality) can increase the probability of resistance development (44, 48). In this context, plants expressing *cry2Aa2* through the chloroplast either singly, or as part of a gene-pyramid with other insect proteins (preferably non-Bt proteins with different modes of action) could become an invaluable tool for resistance management.

Engineering Pathogen Resistance via the Chloroplast Genome

Since the beginning of civilization, plant diseases have plagued global crop production. Between 1979 and 1980 India lost up to 60% of its rice crop due to bacterial rice blight. Between 1988 and 1990, there was a 10.1% loss of the global barley crop due to bacterial pathogens, worth \$1.9 billion (50). In the United States, there was an estimated 44,600 metric ton reduction of soybean crops due to bacterial pathogens in 1994. Many efforts have been made to combat these devastating pathogens. Plant breeding was introduced to fight plant diseases (49). However, results were limited due to the ability of the bacteria to adapt and find a way around the defense mechanism.

Agrochemicals have been used but their application is limited by their toxicity to humans and the environment (49). With the emergence of molecular biology, researchers have been able to elucidate many of the pathways and products in the plant response to phytopathogens.

In general, the plant defense response can be divided into 3 major categories, early defense (fast), local defense (fast/intermediate) and systemic defense (intermediate to slow, 49). During the early stage, the plant cell is stimulated by contact with pathogen-produced elicitors. Bacterial genes such as *hrp* (hypersensitive response and pathogenicity) or *avr* (avirulence) genes stimulate the plant defense mechanism (50). The most prominent early defense response is the HR (hypersensitive response), which leads to cellular death reducing further infection by the pathogen. Local defense entails cell wall reinforcement, stimulation of secondary metabolite pathways, synthesis of thionins and synthesis of PR (pathogenesis-related) proteins (49). The final phase is known as SAR (systemic acquired resistance), which protects the uninfected regions of the plant. Genetic engineering has allowed for the enhancement of natural defense genes from plants by cloning and over expression in non-host plants. Cloning of resistance (R) genes has been used to protect rice from bacterial leaf blight (49). Pathogenesis-related (PR) genes have been cloned from barley and have shown to provide resistance to *P. syringae* pv. *Tabaci* (49). Anti-fungal peptides produced by various organisms have been cloned and studied. While progress made to date is promising for anti-fungal activity (51), bacteria still maintain the ability to adapt to plant defenses.

Plants are not the only species to have problems with pathogenic bacteria. It is common knowledge that the medical community has been fighting a losing battle against pathogenic bacteria for years. The number of multiple drug resistant strains of bacteria is growing, reducing the available choices of antibiotic that can be used. Research continues on multiple fronts to combat antibiotic resistance (52). Cationic antibacterial peptides from mammals, amphibians and insects have gained more attention over the last decade (53). Key features of these cationic peptides are a net positive charge, an affinity for negatively-charged prokaryotic membrane phospholipids over neutral-charged eukaryotic membranes and the ability to form aggregates that disrupt the bacterial membrane (54). Given the fact that the outer membrane is an essential and highly conserved part of all bacterial cells, it would seem highly unlikely that bacteria would be able to adapt (as they have against antibiotics) to resist the lytic activity of these peptides.

There are three major peptides with α -helical structures, cecropin from *Hyalophora cecropia* (giant silk moth), magainins from *Xenopus laevis* (African frog) and defensins from mammalian neutrophils. Magainin and its analogues have been studied as a broad-spectrum topical agent, a systemic antibiotic; a wound-healing stimulant; and an anticancer agent (55). However, the possible agricultural use of magainin-type antimicrobial peptides has not yet been explored. We have recently observed that a

synthetic lytic peptide (MSI-99) can be successfully expressed in tobacco chloroplast (17). The peptide retained its lytic activity against the phytopathogenic bacteria *Pseudomonas syringae* and multidrug resistant human pathogen, *Pseudomonas aeruginosa*. The anti-microbial peptide (AMP) used in this study was an amphipathic alpha-helix molecule that has an affinity for negatively charged phospholipids commonly found in the outer-membrane of bacteria. Upon contact with these membranes, individual peptides aggregate to form pores in the membrane, resulting in bacterial lysis. Because of the concentration dependent action of the AMP, it was expressed via the chloroplast genome to accomplish high dose delivery at the point of infection. PCR products and Southern blots confirmed chloroplast integration of the foreign genes and homoplasmy. Growth and development of the transgenic plants was unaffected by hyper-expression of the AMP within chloroplasts. *In vitro* assays with T₀ and T₁ plants confirmed that the AMP was expressed at high levels (21.5 to 43% of the total soluble protein) and retained biological activity against *Pseudomonas syringae*, a major plant pathogen. *In situ* assays resulted in intense areas of necrosis around the point of infection in control leaves, while transformed leaves showed no signs of necrosis (200-800 µg of AMP at the site of infection). T₁ *in vitro* assays against *Pseudomonas aeruginosa* (a multi-drug resistant human pathogen) displayed a 96% inhibition of growth. These results give a new option in the battle against phytopathogenic and drug-resistant human pathogenic bacteria.

Engineering Novel Pathways via the Chloroplast Genome

We have recently used the *Bacillus thuringiensis* (Bt) *cry2Aa2* operon as a model system to demonstrate operon expression and crystal formation via the chloroplast genome (12). *Cry2Aa2* is the distal gene of a three-gene operon. The *orf* immediately upstream of *cry2Aa2* codes for a putative chaperonin that facilitates the folding of *cry2Aa2* (and other proteins) to form proteolytically stable cuboidal crystals (56). Because CRY protein levels decrease in plant tissues late in the growing season or under physiological stress (57), a more stable protein expressed at high levels in the chloroplast throughout the growing season should increase toxicity of Bt transgenic plants to target insects and help eliminate the development of Bt resistance.

Therefore, the *cry2Aa2* bacterial operon was expressed in tobacco chloroplasts to test the resultant transgenic plants for increased expression and improved persistence of the accumulated insecticidal protein(s). Stable foreign gene integration was confirmed by PCR and Southern blot analysis in T₀ and T₁ transgenic plants. *Cry2Aa2* operon derived protein accumulated at 45.3% of the total soluble protein in mature leaves and remained stable even in old bleached leaves (46.1%). This is the highest level of foreign gene expression ever reported in transgenic plants. Exceedingly difficult to control insects (10-day old cotton boll worm, beetarmy worm) were killed 100% after

consuming transgenic leaves. Electron micrographs showed the presence of the insecticidal protein folded into cuboidal crystals similar in shape to Cry2Aa2 crystals observed in *Bacillus thuringiensis*. In contrast to currently marketed transgenic plants with soluble CRY proteins, folded protoxin crystals will be processed only by target insects that have alkaline gut pH; this approach should improve efficacy of Bt transgenic plants. Absence of insecticidal proteins in transgenic pollen eliminates toxicity to non-target insects via pollen. In addition to these environmentally friendly approaches, this observation should serve as a model system for large-scale production of foreign proteins within chloroplasts in a folded configuration enhancing their stability and facilitating single step purification. This is the first demonstration of expression of a bacterial operon in transgenic plants and opens the door to engineer novel pathways in plants in a single transformation event. Since then, we have engineered the *mer* operon via the chloroplast genome to study phytoremediation.

Engineering Abiotic Stress Tolerance via the Chloroplast Genome

Water stress due to drought, salinity or freezing is a major limiting factor in plant growth and development. Trehalose is a non-reducing disaccharide of glucose and its synthesis is mediated by the trehalose-6-phosphate (T6P) synthase and trehalose-6-phosphate phosphatase complex in *Saccharomyces cerevisiae*. In *S. cerevisiae*, this complex consists of at least three subunits performing either T6P synthase (*TPS1*), T6P phosphatase (*TPS2*) or regulatory activities (*TPS3* or *TSL1*, 58, 59). Trehalose is found in diverse organisms including algae, bacteria, insects, yeast, fungi, animal and plants (60). Because of its accumulation under various stress conditions such as freezing, heat, salt or drought, there is general consensus that trehalose protects against damages imposed by these stresses (61-63). Trehalose is also known to accumulate in anhydrobiotic organisms that survive complete dehydration (64), the resurrection plant (65) and some desiccation tolerant angiosperms (66). Trehalose, even when present in low concentrations, stabilizes proteins and membrane structures under stress (67) because of the glass transition temperature, greater flexibility and chemical stability/inertness.

As pointed out earlier, chloroplast transformation has several other advantages over nuclear transformation (1). The difficulty in accomplishing gene containment in nuclear transgenic plants is a serious concern, especially when plants are genetically engineered for drought tolerance, because of the possibility of creating robust drought tolerant weeds and passing on undesired pleiotropic traits to related crops. Chloroplast transformation should also overcome some of the disadvantages of nuclear transformation that result in lower levels of foreign gene expression, such as gene suppression by positional effect or gene silencing (18,19). Therefore, we have recently

introduced the yeast *trehalose phosphate synthase* (*TPS1*) gene into the tobacco chloroplast and nuclear genomes to study resultant phenotypes (28). PCR and Southern blots confirmed stable integration of *TPS1* into the chloroplast genomes of T₁, T₂ and T₃ transgenic plants. Northern blot analysis of transgenic plants showed that the chloroplast transformant expressed 16,966-fold more *TPS1* transcript than the best surviving nuclear transgenic plant. Although both the chloroplast and nuclear transgenic plants showed significant *TPS1* enzyme activity, no significant trehalose accumulation was observed in T₀/T₁ nuclear transgenic plants whereas chloroplast transgenic plants showed 15-25 fold higher accumulation of trehalose than the best surviving nuclear transgenic plants. Nuclear transgenic plants (T₀) that showed significant amounts of trehalose accumulation showed stunted phenotype, sterility and other pleiotropic effects whereas chloroplast transgenic plants (T₁, T₂, T₃) showed normal growth and no pleiotropic effects. Chloroplast transgenic plants also showed a high degree of drought tolerance as evidenced by growth of transgenic plants in 6% polyethylene glycol whereas respective control plants were bleached. After 7hr air drying, chloroplast transgenic plants (T₁, T₂, T₃) successfully rehydrated while control plants died. In order to prevent escape of drought tolerance trait to weeds and associated pleiotropic traits to related crops, it is desirable to genetically engineer crop plants for drought tolerance via the chloroplast genome instead of the nuclear genome.

Marker Free Transgenic Plants Engineered via the Chloroplast Genome

Despite several advantages, one major disadvantage with chloroplast genetic engineering may be the utilization of the antibiotic resistance genes as the selectable marker to confer streptomycin/spectinomycin resistance. When this selection process for chloroplast genetic engineering was first investigated, the mutant 16S rRNA gene that does not bind the antibiotic was used (68). Subsequently, the *aadA* gene product that inactivates the antibiotic by transferring the adenyl moiety of ATP to spectinomycin /streptomycin was used (69). These antibiotics are commonly used to control bacterial infection in humans and animals. The probability of gene transfer from plants to bacteria living in the gastrointestinal tract or soil may be enhanced by the compatible protein synthetic machinery between chloroplasts and bacteria, in addition to presence of thousands of copies of the antibiotic resistance genes per cell. Also, most antibiotic resistance genes used in genetic engineering originate from bacteria. Therefore, betaine aldehyde dehydrogenase (BADH) gene from spinach is used in this study as a selectable marker. The selection process involves conversion of toxic betaine aldehyde (BA) by the chloroplast BADH enzyme to nontoxic glycine betaine, which also serves as an

osmoprotectant (70). This enzyme is present only in chloroplasts of a few plant species adapted to dry and saline environments (70, 29).

Daniell et al (71) report development of an antibiotic free selection process for chloroplast genetic engineering, utilizing a gene that is naturally present in spinach. Chloroplast transformation efficiency was 25 fold higher in BA selection than spectinomycin, in addition to rapid regeneration. Transgenic shoots appeared within 12 days in 80% of leaf discs (up to 23 shoots per disc) in BA selection compared to 45 days in 15% of discs (1 or 2 shoots per disc) on spectinomycin selection. Southern blots confirm stable integration of foreign genes into all of the chloroplast genomes (~10,000 copies per cell) resulting in homoplasmy. Transgenic tobacco plants showed 1527-1816% higher BADH activity at different developmental stages than untransformed controls. Transgenic plants were morphologically indistinguishable from untransformed plants and the introduced trait was stably inherited in the subsequent generation. This is the first report of genetic engineering of the chloroplast genome without the use of antibiotic resistance genes. Use of genes that are naturally present in spinach for selection, in addition to gene containment, should ease public concerns or perception of GM crops.

Production of Pharmaceutical Proteins via Chloroplast Genome

Research on human proteins in the past years has revolutionized the use of these therapeutically valuable proteins in a variety of clinical situations. Since the demand for these proteins is expected to increase considerably in the coming years, it would be wise to ensure that in the future they will be available in significantly larger amounts, preferably on a cost-effective basis. Because most genes can be expressed in many different systems, it is essential to determine which system offers the most advantages for the manufacture of the recombinant protein. The ideal expression system would be one that produces a maximum amount of safe, biologically active material at a minimum cost. The use of modified mammalian cells with recombinant DNA techniques has the advantage of resulting in products which are closely related to those of natural origin; however, culturing of these cells is intricate and can only be carried out on limited scale. The use of microorganisms such as bacteria permits manufacture on a larger scale, but introduces the disadvantage of producing products, which differ appreciably from the products of natural origin. For example, proteins that are usually glycosylated in humans are not glycosylated by bacteria. Furthermore, human proteins that are expressed at high levels in *E. coli* frequently acquire an unnatural conformation, accompanied by intracellular precipitation due to lack of proper folding and disulfide bridges. Production of recombinant proteins in plants has many potential advantages for generating biopharmaceuticals relevant to clinical medicine. These include the following: (I) plant systems are more

economical than industrial facilities using fermentation systems; (ii) technology is available for harvesting and processing plants/ plant products on a large scale; (iii) elimination of the purification requirement when the plant tissue containing the recombinant protein is used as a food (edible vaccines); (iv) plants can be directed to target proteins into stable, intracellular compartments as chloroplasts, or expressed directly in chloroplasts; (v) the amount of recombinant product that can be produced approaches industrial-scale levels; and (vi) health risks due to contamination with potential human pathogens/toxins are minimized.

It has been estimated that one tobacco plant should be able to produce more recombinant protein than a 300-liter fermenter of *E. coli*. In addition, a tobacco plant produces a million seeds, facilitating large-scale production. Tobacco is also an ideal choice because of its relative ease of genetic manipulation and an impending need to explore alternate uses for this hazardous crop. However, with the exception of enzymes (e.g. phytase), levels of foreign proteins produced in nuclear transgenic plants are generally low, mostly less than 1% of the total soluble protein (72). Daniell et al. (16) discuss this problem using the following examples. Although plant derived recombinant hepatitis B surface antigen was as effective as a commercial recombinant vaccine, the levels of expression in transgenic tobacco were low (0.0066% of total soluble protein). Even though Norwalk virus capsid protein expressed in potatoes caused oral immunization when consumed as food (edible vaccine), expression levels were low (0.3% of total soluble protein). In particular, expression of human proteins in nuclear transgenic plants has been disappointingly low: e.g. human Interferon- α 0.000017% of fresh weight, human serum albumin 0.02% and erythropoietin 0.0026% of total soluble protein (see table 1 in ref 72). A synthetic gene coding for the human epidermal growth factor was expressed only up to 0.001% of total soluble protein in transgenic tobacco (16). The cost of producing recombinant proteins in alfalfa leaves was estimated to be 12-fold lower than in potato tubers and comparable with seeds (72). However, tobacco leaves are much larger and have much higher biomass than alfalfa. The cost of production of recombinant proteins will be 50-fold lower than that of *E. coli* fermentation (with 20% expression levels, 73). A decrease in insulin expression from 20% to 5% of biomass doubled the cost of production (2b). Expression level less than 1% of total soluble protein in plants has been found to be not commercially feasible (72). Therefore, it is important to increase levels of expression of recombinant proteins in plants in order to exploit plant production of pharmacologically important proteins.

An alternate approach is to express foreign proteins in chloroplasts of higher plants. Most importantly, a significant advantage in the production of pharmaceutical proteins in chloroplasts is their ability to process eukaryotic proteins, including folding and formation of disulfide bridges (74). Chaperonin proteins are present in chloroplasts (75,76) that function in folding and assembly of prokaryotic/eukaryotic proteins. Also, proteins are

activated by disulfide bond oxido/reduction cycles using the chloroplast thioredoxin system (77) or chloroplast protein disulfide isomerase (78). Accumulation of fully assembled, disulfide bonded form of human somatotropin via chloroplast transformation (15) and oligomeric form of CTB (79) and assembly of heavy and light chains of humanized Guy's 13 antibody in transgenic chloroplasts (80) provide strong evidence for successful processing of pharmaceutical proteins inside chloroplasts. Such folding and assembly should eliminate the need for highly expensive *in vitro* processing of pharmaceutical proteins. For example, 60% of the total operating cost in the production of human insulin is associated with *in vitro* processing (formation of disulfide bridges and cleavage of methionine, 73).

Taken together, low levels of expression of human proteins in nuclear transgenic plants, and difficulty in folding, assembly/processing of human proteins in *E.coli* should make chloroplasts an ideal compartment for expression of these proteins; production of human proteins in transgenic chloroplasts should also dramatically lower the production cost. Large-scale production of these proteins in plants should be a powerful approach to provide treatment to patients at an affordable cost and provide tobacco farmers alternate uses for this hazardous crop. Therefore, therapeutic proteins have been expressed in transgenic tobacco chloroplasts to increase levels of expression and accomplish *in vivo* processing. The following are few such examples.

Expression of the Human Somatotropin via the Chloroplast Genome

Staub et al (15) reported a feasibility study for production of a human therapeutic protein through chloroplast transformation technology, which has the additional advantage of increased biological containment by apparent elimination of the transmission of transgenes through pollen. Chloroplasts expressed a secretory protein, human somatotropin, in a soluble, biologically active, disulfide-bonded form. High concentrations of recombinant protein accumulation were observed (>7% total soluble protein), more than 300-fold higher than a similar gene expressed using a nuclear transgenic approach. The plastids-expressed somatotropin was nearly devoid of complex post-translational modifications, effectively increasing the amount of usable recombinant protein. Authors also described approaches to obtain a somatotropin with a non-methionine N terminus, similar to the native human protein. The results indicate that chloroplasts are a highly efficient vehicle for the potential production of pharmaceutical proteins in plants.

Expression of Cholera Toxin β Subunit Oligomers as an Edible Vaccine in Transgenic Chloroplasts:

Vibrio cholerae, which causes acute watery diarrhea by colonizing the small intestine and producing the enterotoxin, cholera toxin (CT). Cholera toxin is a hexameric AB₅ protein consisting of one toxic 27kDa A subunit having ADP ribosyl transferase activity and a nontoxic pentamer of 11.6 kDa B subunits (CTB) that binds to the A subunit and facilitates its entry into the intestinal epithelial cells. CTB when administered orally is a potent mucosal immunogen which can neutralize the toxicity of the CT holotoxin by preventing it from binding to the intestinal cells. This is believed to be a result of it binding to eukaryotic cell surfaces via the GM₁ gangliosides, receptors present on the intestinal epithelial surface, thus eliciting a mucosal immune response to pathogens (81) and enhancing the immune response when chemically coupled to other antigens (82-85).

Cholera toxin (CTB) has previously been expressed in nuclear transgenic plants at levels of 0.01 (leaves) to 0.3% (tubers) of the total soluble protein. To increase expression levels, we engineered the chloroplast genome to express the CTB gene (82). We observed expression of oligomeric CTB at levels of 4-5% of total soluble plant protein. PCR and Southern Blot analyses confirmed stable integration of the CTB gene into the chloroplast genome. Western blot analysis showed that transgenic chloroplast expressed CTB was antigenically identical to commercially available purified CTB antigen. Also, GM₁-ganglioside binding assays confirm that chloroplast synthesized CTB binds to the intestinal membrane receptor of cholera toxin. Transgenic tobacco plants were morphologically indistinguishable from untransformed plants and the introduced gene was found to be stably inherited in the subsequent generation as confirmed by PCR and Southern Blot analyses. The increased production of an efficient transmucosal carrier molecule and delivery system, like CTB, in chloroplasts of plants makes plant based oral vaccines and fusion proteins with CTB needing oral administration, a much more feasible approach. This also establishes unequivocally that chloroplasts are capable of forming disulfide bridges to assemble foreign proteins.

Expression and Assembly of Monoclonals in Transgenic Chloroplasts:

Dental caries (cavities) is probably the most prevalent disease of humankind. Colonization of teeth by *S. mutans* is the single most important risk factor in the development of dental caries. *S. mutans* is a non-motile, gram positive coccus. It colonizes tooth surfaces and synthesizes glucans (insoluble polysaccharide) and fructans from sucrose using the enzymes glucosyltransferase and fructosyltransferase respectively (86). The glucans play an important role by allowing the bacterium to adhere to the smooth

tooth surfaces. After its adherence, the bacterium ferments sucrose and produces lactic acid. Lactic acid dissolves the minerals of the tooth, producing a cavity.

A topical monoclonal antibody therapy to prevent adherence of *S. mutans* to teeth has recently been developed. The incidence of cariogenic bacteria (in humans and animals) and dental caries (in animals) was dramatically reduced for periods of up to two years after the cessation of the antibody therapy. No adverse events were detected either in the exposed animals or in human volunteers (87). The annual requirement for this antibody in the US alone may eventually exceed 1 metric ton. Therefore, this antibody was expressed via the chloroplast genome to achieve higher levels of expression and proper folding (84). The integration of antibody genes into the chloroplast genome was confirmed by PCR and Southern blot analysis. The expression of both heavy and light chains was confirmed by western blot analysis under reducing conditions. The expression of fully assembled antibody was confirmed by western blot analysis under non-reducing conditions. This is the first report of successful assembly of a multi-subunit human protein in transgenic chloroplasts. Production of monoclonal antibodies at agricultural level should reduce their cost and create new applications of monoclonal antibodies.

Challenges Facing Chloroplast Genetic Engineering

While there are several reports of genetic engineering of the chloroplast genome in tobacco, other major crops (including cereals) have not yet been exploited. One of the major limitations has been the lack of knowledge of chloroplast genome sequences to locate spacer regions and transcriptional units to target site-specific integration of foreign genes. In order to overcome this limitation, Daniell et al. (6) have recently developed a universal vector that can transform any chloroplast genome because it integrates into a highly conserved region. Because of the conservation of plastid genome sequence across many plant species, the strategy of using targeting sequences from one species to transform the plastid genome of another unknown species was developed (6). Both potato (88) and more recently tomato (89) were transformed using chloroplast vectors with tobacco targeting sequences and tobacco was transformed with petunia targeting sequences (17). Experiments are in progress to transform a variety of crops using this universal vector.

Another limitation has been the ability to regenerate plants only from embryonic tissues in cereals and not from mesophyll cells. Cells from embryogenic tissues contain only proplastids and not mature plastids. It has been suggested that these plastids are smaller than the size of microprojectiles used for DNA delivery and therefore may pose problems in transformation experiments. Successful expression of chloramphenicol acetyl transferase in

proplastids of NT1 cells (90) and β -glucuronidase in proplastids of wheat embryos (91) via particle bombardment suggest that particle size may not be a problem in transforming proplastids. Some of the challenges in transforming agronomically useful crops include optimization of tissue culture techniques and the selection process to obtain transgenic plants via particle bombardment, especially from non-green tissues. Even if homoplasmy is not obtained in the first generation, it could be accomplished in subsequent generations by germination of T1 seeds under appropriate selection. In this context, it should be noted that the use of chloroplast integration vectors with ori sequences has been shown to accomplish homoplasmy even in the first round of selection (14). Also, heteroplasmy of chloroplast genomes has been observed in nature (92) and accomplishing homoplasmy for the introduced trait may not be always necessary.

Yet another concern is the possibility of yield drag in transgenic crops because of the hyper-expression of foreign genes via the chloroplast genome (at times as high as 30% of the total soluble protein). High levels of expression of several foreign proteins in transgenic tobacco has not affected growth rates, photosynthesis, chlorophyll content, flowering, or seed setting (12). Chloroplasts are used to handling such abundant proteins without deleterious effect on productivity. For example, the Calvin cycle enzyme, ribulose bis-phosphate carboxylase/oxygenase (RubisCO) is synthesized as much as 50% of the total soluble protein; such high levels of synthesis have not affected the productivity of crop plants. Indeed, excess RubisCO is constantly made and degraded in chloroplasts. However, long term tests using agronomically important crops grown under field conditions are needed to confirm this observation. Recent success in accomplishing potato and tomato plastid transformation should pave the way for studies on such agronomic traits. All of these findings augur well for chloroplast genetic engineering of economically useful crops. Thus several environmentally friendly approaches have been opened for new advances in plant biotechnology and genetic engineering.

REFERENCES

1. Bogorad, L. (2000) Engineering chloroplasts: an alternative site for foreign genes, Proteins, reactions and products. *Trends in Biotechnology* **18**: 257-263.
2. Daniell, H. (1999) Environmentally friendly approaches to genetic engineering. In *Vitro Cellular and Developmental Biology-Plant*. **35**: 361-368.
3. Keeler, K.H., Tumer, C.E., and Bolick, M.R. (1996) Movement of crop transgenes into wild plants. In *Herbicide Resistant Crops*, Duke (eds.). CRC Press, pp. 303-330.
4. Hoyle, B. (1999) Canadian farmers seek compensation for genetic pollution. *Nature Biotechnol.* **17**: 747-748.
5. Fox, J.L. (2000) GM food singled out for labeling in the U.S. *Nature Biotechnol.* **18**: 375.
6. Daniell, H., Datta, R., Varma, S., Gray, S., LEE, S.B. (1998) Containment of herbicide resistance through genetic engineering of the chloroplast genome. *Nature Biotechnol.* **16**: 345-348.
7. Scott, S.E., and Wilkinson, M.J. (1999) Risks of transgene escape from transplastomic oilseed rape. *Nature Biotechnology*, **17**: 390-392.

8. Daniell, H. (2000) Genetically modified food crops: current concerns and solutions for the next generation crops. *Biotechnology and Genetic Engineering Reviews*, **17**: 327-347.
9. Kota, M., Daniell, H., Varna, S., Garczynski, F., Gould, F., and Moar, W.J. (1999) Overexpression of the *Bacillus thuringiensis* Cry2A protein in chloroplasts confers resistance to plants against susceptible and Bt-resistant insects. *Proc. Natl. Acad. Sci. USA*. **96**: 1840-1845.
10. Losey, J.E., Rayor, L.S., and Carter, M.C. (1999) Transgenic pollen harms monarch larvae. *Nature*. **399**: 214.
11. Hodgson, J. (1999) Monarch Bt-corn paper questioned. *Nature Biotechnol.* **17**: 627.
12. DeCosa, B., Moar, W., Lee, S.B., Miller, M., Daniell, H. (2001), *Nature Biotechnology*, **19**: 71-74.
13. Daniell, H. (1999) New tools for chloroplast genetic engineering. *Nature Biotechnol.* **17**: 855-856.
14. Guda, C., Lee, S.B., and Daniell, H. (2000) Stable expression of biodegradable protein based polymer in tobacco chloroplasts. *Plant Cell Rep.* **19**: 257-262.
15. Staub J.M., Garcia B., Graves J., Hajdukiewicz P.T., Hunter P., Nehra N., Paradkar V., Schlittler M., Carroll J.A., Spatola L., Ward D., Ye G., and Russell D.A. (2000) High-yield production of a human therapeutic protein in tobacco chloroplasts. *Nat. Biotechnol.* **18**:333-338.
16. Daniell, H. Streatfield, S., Wycoff, K.. (2001). Medical molecular farming: production of antibodies, biopharmaceuticals and edible vaccines in plants. *Trends Plant Sci.* **6**, 219-26
17. GeDray, G., Smith, F., Sanford, J., and Daniell, H. (2000), Hyper-expression of of an antimicrobial peptide via the chloroplast genome to confer resistance against phytopathogenic bacteria and fungi. *Plant Physiology* **127**: 1-11.
18. Vaucheret, H. et al. (1998), Transgene induced gene silencing in plants. *Plant J.* **16**, 651-659.
19. De Neve, M. et al. (1999), Gene silencing results in instability of antibody production in transgenic plants. *Mol. Gen. Genetics* **260**, 582-592.
20. Navrath, C., Poirier, Y., and Somerville, C. (1994) Targeting of the polyhydroxy butyrate biosynthetic pathway to the plastids of *Arabidopsis thaliana* results in high levels of polymer accumulation. *Proc. Natl. Acad. Sci.* **91**, 12760-12764.
21. Ma, J. et al. (1995) Generation and assembly of secretory antibodies in plants *Science*. **268**, 716-719.
22. Ye, X. et al. (2000) Engineering the provitamin A (β -carotene) biosynthetic pathway into (carotenoid-free) rice endosperm. *Science*. **287**, 303-305.
23. Holmstrom, K.O., Mantyla, M., Wekin, B., Mandal, A., Palva, E.T., Tunnela, O.E., and Londesborough, J. (1996) Drought tolerance in tobacco. *Nature*. **379**, 683-684.
24. Goddijn, O.J.M., Verwoerd, T.C., Voogd, E., Krutwagen, W.H.H., de Graff, P.T.H.M., Poels, J., van Dun, K., Ponstein, A.S., Damm, B., and Pen, K. (1997) Inhibition of trehalase activity enhances trehalose accumulation in transgenic plants. *Plant Physiol.* **113**, 181-190.
25. Romero, C., Belles, J.M., Vaya, J.L., Serrano, R., and Culianz-Macia, F.A. (1997) Expression of the yeast trehalose-6-phosphate synthase gene in transgenic tobacco plants: pleiotropic phenotypes include drought tolerance. *Planta*. **201**, 293-297.
26. During, K., Hippe, S., Kreuzaler, F., and Schell, J. (1990) Synthesis and self-assembly of a functional monoclonal antibody in transgenic *Nicotiana tabacum*. *Plant Molecular Biology*. **15**, 281-293 (1999).
27. Daniell, H. and Guda, C. (1997) Biopolymer production in microorganisms and plants. *Chemistry and industry*. **14**, 555-560.
28. Lee, S.B., Lee SB, Kwon H, Kwon S, Park S, Jeong M, Han S, Byun H., Daniell, H (2001). Drought tolerance conferred by the yeast trehalose-6 phosphate synthase gene engineered via the chloroplast genome. *Transgenic Research*, in press.
29. Nuccio, M.L., Rhodes, D., McNeil, S.D., and Hanson, A.D. (1999) Metabolic engineering of plants for osmotic stress tolerance. *Curr. Opinion in Plant Biology* **2**:128-134.
30. Puchta, H. (2000) Removing selectable marker genes: taking the shortcut. *Trends in Plant*

Science **5**: 273-274.

31. Daniell, H. (1999) GM crops: Public perception and scientific solutions. *Trends in Plant Science* **4**: 467-469.
32. Peerenboom, E. (2000) German health minister calls time out for Bt maize. *Nature Biotechnol.* **18**: 374.
33. Shah, D.M., Horsch, R.B., Klee, H.J., Kishore, G.M., Winter, J.A., Turner, E.N., Hironaka, C.M., Sanders, P.R., Gasser, C.S., Aykent, S., Siegel, N.R., Rogers, S.G., and Fraley, R.T. (1986) Engineering herbicides tolerance in transgenic plants. *Science* **233**, 478-481.
34. Cioppa, G.D., Baner, S.C., Tayler, M.L., Roshester, D.E., Klein, B.K., Shah, D. M., Fraley, R.T., and Kishore, G.M. (1987) Targeting a herbicide resistant enzyme from *E. coli* to chloroplasts of higher plants. *Bio/Technology* **5**, 579-584.
35. Gressel, J. 1999, Tandem constructs: preventing the rise of superweeds. *TIBTECH* **17**, 361-366.
36. Mariani, C., DeBeuckeleer, M., Trueltner, J, Leemans, J., and Goldberg, R.B. (1990) Induction of male sterility in plants by a chimaeric ribonuclease gene. *Nature* **347**:737-741.
37. Daniell, H. , and Varma, S. (1998), Chloroplast transgenic plants: Panacea-No. Gene Containment-Yes. *Nature Biotechnology* **16**: 602
38. Maier, R.M., Neckermann, K., Igloi, G.L., and Kössel, H. (1995) Complete sequence of the maize chloroplast genome: gene content, hotspots of divergence and fine tuning of genetic information by transcript editing. *J. Mol. Biol.* **251**, 614-628.
39. Daniell, H. (1993) Foreign gene expression in chloroplasts of higher plants mediated by tungsten particle bombardment. *Methods in Enzymology.* **217**: 536-556.
40. Daniell, H. (1997) Transformation and foreign gene expression in plants mediated by microprojectile bombardment. *Meth. Mol. Biol.* **62**: 453-488.
41. Tabashnik, B.E., Cushing, N.L., Finson, N., and Johnson, M.W. (1990) Field development of resistance to *Bacillus thuringiensis* in diamond back moth. *J. Econ. Entomol.* **83**, 1671-1676.
42. Koziel, M.G., Beland, G.L., Bowman, C., Carozzi, N.B., Crenshaw, R., Crossland, L., Dawson, J., Desai, N., Hill, M., Kadwell, S., Launis, K., Lewis, K., Maddox, D., McPherson, K., Meghji, M.R., Merlin, E., Rhodes, R., Warren, G.W., Wright, M., and Evola, S.V. (1993) FIELD PERFORMANCE OF ELITE transgenic maize plants expressing an insecticidal protein derived from *Bacillus thuringiensis*. *Bio/Technol.* **11**, 194-200.
43. Perlak, F.J., Deaton, R.W., Armstrong, T.A., Fuchs, R.L., Sims, S.R., Greenplate, J.T., and Fischhoff, D.A., (1990) Insect resistant cotton plants. *Bio/Technol.* **8**, 939-943.
44. Gould, F. (1998) Sustainability of transgenic insecticidal cultivars: integrating pest genetics and ecology. *ANNU. Rev. Entomol.* **43**, 701-726.
45. Gould, F., Martinez-Ramirez, A., Ferre, J., Silva, F.J., and Moar, W. (1992) Broad spectrum resistance to *Bacillus thuringiensis* toxins in *Heliothis virescens*. *Proc. Natl. Acad. Sci. USA* **89**, 7986-7990.
46. Gould, F., Anderson, A., Reynolds, A., Bumgarner, L., and Moar, W. (1995) Selection and genetic analysis of a *Heliothis virescens* strain with high levels of resistance to *Bacillus thuringiensis* toxins. *J. Econ Entomol.* **88**, 1545-1559.
47. Liu, Y.B., Tabashnik, B.E., Dennehy, T.J., Patin, A.L., and Barlett, A.C. (1999) Development time and resistance to Bt crops. *Nature* **400**, 519.
48. Tabashnik, B.E., Liu, Y.B., Finson, N., Masson, L., and Heckel, D.G. (1997) One gene in diamond back moth confers resistance to four *Bacillus thuringiensis*. *Proc. Natl. Acad. Sci. USA* **94**:1640-1644.
49. Mourgues, F., Brisset, M.N., and Cheveau, E. (1998), Strategies to improve plant resistance to bacterial diseases through genetic engineering. *Trends in Biotechnology* **16**, 203-210.
50. Baker, B., Zambryski, P., Staskawicz, S., and Dinesh-Kumar, P. (1997), Signalling in plant-micobe interactions. *Science* **276**, 723-726.

51. Cary et al. 2000, Transgenic expression of a gene coding a synthetic antimicrobial peptide results in inhibition of fungal growth in vitro and in planta. *Plant Science* **154**, 171-181.
52. Persidis, A. (1999), Antibacterial and antifungal drug discovery. *Nature Biotechnology* **17**, 1141-1142.
53. Hancock, R., Lehrer, R. (1998), Cationic peptides: a new source of antibiotics. *TIBTECH* **16**, 82-88.
54. Biggin, P., Sansom, M. 1999, Interactions of α -helices with lipid bilayers: a review of simulation studies. *Biophysical Chemistry* **76**, 161-183.
55. Jacob, L., Zasloff, M. (1994), Potential therapeutic applications of megainins and other antimicrobial agents of animal origin. *Ciba Foundation Symposium* **186**, 197-223.
56. Ge, B. et al. (1998) Differential effects of helper proteins encoded by the *cry2A* and *cry11A* operons on the formation of Cry2A inclusions in *Bacillus thuringiensis*. *FEMS Microbiol. Lett.* **165**, 35-41.
57. Greenplate, J. (1999), Quantification of *Bacillus thuringiensis* insect control protein Cry1Ac over time in bollgard cotton fruit and terminals. *J. Econ. Entomol.* **92**, 1377-1383.
58. Thevelein, J.M. and Hohmann, S. (1995), Trehalose synthase: guard to the gate of glycolysis in yeast? *Trends in Bioscience.* **20**, 3-10.
59. Singer, M.A. and Lindquist, S. (1998), Thermotolerance in *Saccharomyces cerevisiae*: the Yin and Yang of trehalose. *Trends in Biotech.* **16**, 460-468.
60. Elbein, A.D. (1974) The metabolism of α , α -trehalose. *Adv Carbohydr Chem Biochem.* **30**, 227-256.
61. Mackenzie, K.F., Singh, K.K., and Brown, A.D. (1988) Water stress plating hypersensitivity of yeast: protective role of trehalose in *Saccharomyces cerevisiae*. *J Gen Microbial.* **134**, 1661-1666.
62. De Vigilio, C., Hottinger, T., Dominguez, J., Boller, T., and Wickman, A. (1994) The role of trehalose synthesis for the acquisition of thermotolerance in yeast I. Genetic evidence that trehalose is a thermoprotectant. *Eur J Biochem.* **219**, 179-186.
63. Sharma, S.C. (1997) A possible role of trehalose in osmotolerance and ethanol tolerance in *Saccharomyces cerevisiae* *Fems Microbiology Letters.* **152**, 11-15.
64. Crowe J.H., Hoekstra F.A. and Crowe L.M. (1992) Anhydrobiosis. *Annu Rev Physiol.* **54**, 579-599.
65. Bianchi, G., Gamba, A., and Limiroli, R. (1993) The unusual sugar composition in leaves of the resurrection plant *Myrothamnus flabellifolia*. *Physiol Plantarum.* **87**, 223-226.
66. Drennan, P.M., Smith, M.T., Goldsworthy, D., and Van Staden, J. (1993) The occurrence of trehalose in the leaves of the desiccation-tolerant angiosperm *Myrothamnus flabellifolius* Welw. *J. Plant Physiol.* **142**, 493-496.
67. Iwahashi, H., Obuchi, K., Fujii, S., and Komatsu, Y. (1995) The correlative evidence suggesting that trehalose stabilizes membrane-structure in the yeast *Saccharomyces cerevisiae*. *Cell. Mol. Biol.* **41**, 763-769.
68. Svab, Z., Hajdukiewicz P., and Maliga P (1990) Stable transformation of plastids in higher plants. *Proc. Natl. Acad. Sci. USA* **87**: 8526-8530.
69. Svab Z., Maliga P (1993) High frequency plastid transformation in tobacco by selection for a chimeric *aadA* gene. *Proc. Natl. Acad. Sci. USA.* **90**: 913-917.
70. Rathinasabapathy, B., McCue, K.F., Gage, D.A., and Hanson, A.D. (1994) Metabolic engineering of glycine betaine synthesis: Plant betaine aldehyde dehydrogenases lacking typical transit peptides are targeted to tobacco chloroplasts where they confer aldehyde resistance. *Planta.* **193**: 155-162.
71. Daniell, H., Muthukumar, S., and Lee, S.B. (2000), Engineering the chloroplast genome without the use of antibiotic resistance genes. *Current Genetics*, in press

72. Kusnadi A., Nikolov Z., and Howard J. (1997). Production of Recombinant proteins in Transgenic plants: Practical considerations. *Biotechnology and Bioengineering*. **56** (5): 473-484.
73. Petridis D., Sapidou E., and Calandranis J. (1995) Computer aided process analysis and economic evaluation for biosynthetic human insulin production- A case study. *Biotechnology and Bioengineering* **48**: 529-541.
74. Drescher D.F., Follmann H., and Haberlein I. (1998). Sulfitolysis and thioredoxin-dependent reduction reveal the presence of a structural disulfide bridge in spinach chloroplast fructose-1, 6-bisphosphate. *FEBS Letters* **424**: 109-112.
75. Roy H. (1989). Rubisco assembly: a model system for studying the mechanism of chaperonin action. *Plant Cell*. **1**: 1035-1042.
76. Vierling E. (1991). The roles of heat shock proteins in plants. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **42**: 579-620.
77. Reulland E. and Miginiac-Maslow M (1999). Regulation of chloroplast enzyme activities by thioredoxins: activation or relief from inhibition. *Trends in Plant Science* **4**: 136-141.
78. Kim J. and Mayfield P.S. (1997). Protein disulfide isomerase as a regulator of chloroplast translational activation. *Science* **278**: 1954-1957.
79. Daniell H., Lee, S.B., Panchal, T., wiebe, P.O. (2001) Expression of cholera toxin B subunit gene and assembly of functional oligomers in transgenic tobacco chloroplasts. *J. Mol. Biol.* **311**: 1001-1009.
80. Daniell, H., Dhingra, A. & Fernandez-San Millan, A. (2001) Chloroplast transgenic approach for the production of antibodies, biopharmaceuticals and edible vaccines. Proc. Int. Cong. Photosynth. Brisbane, S40-04, Australia.
81. Lipscombe M., Charles I.G., Roberts M., Dougan G., Tite J., and Fairweather N.F. (1991). Intranasal immunization using the B subunit of the *Escherichia coli* heat-labile toxin fused to an epitope of the Bordetella pertussis P.69 antigen. *Mol. Microbiol.* **5**: 1385-1392.
82. Dertzbaugh M.T. and Elson C.O. (1993). Comparative effectiveness of the cholera toxin B subunit and alkaline phosphatase as carriers for oral vaccines. *Infect. Immun.* **61**: 48-55.
83. Holmgren J., Lycke N., and Czerkinsky C. (1993). Cholera toxin and cholera B subunit as oral-mucosal adjuvant and antigen vector systems. *Vaccine*. **11**: 1179-84.
83. Nashar T.O., Amin T., Marcello A., and Hirst T.R. (1993). Current progress in the development of the B subunits of cholera toxin and *Escherichia coli* heat-labile enterotoxin as carriers for the oral delivery of heterologous antigens and epitopes. *Vaccine*. **11**(2): 235-40.
85. Sun J.B., Holmgren J., and Czerkinsky C. (1994). Cholera toxin B subunit: an efficient transmucosal carrier-delivery system for induction of peripheral immunological tolerance. *Proc. Natl. Acad. Sci. USA* **91**: 10795-10799.
86. Hotz P., Guggenheim B., and Schmid R. (1972). Carbohydrates in pooled dental plaque. *Caries Res.* **6**: 103-21.
87. Ma J., Hikmat B., Wycoff K., Vine N., Chargelegue D., Yu L., Hein M., and Lehner T. (1998) Characterization of a recombinant plant monoclonal secretory antibody and preventive immunotherapy in humans. *Nat. Med.* **4**: 601-606.
88. Sidorov, V.A., Kasten, D., Pang, S.G., Hajdukiewicz, P.T.J., Staub, J.M., and Nehra, N.S. (1999) Potato plastid transformation: high expression of green fluorescent protein in chloroplasts. *Plant J.*, **19**, 209-216.
89. Ruf, S., Hermann, M., Berger, I.J., Carrer, H., Bock, R. (2001) Stable genetic transformation of tomato plastids: high level foreign protein expression in fruits. *Nature Biotechnol.* **19**, 870-875.
90. Daniell, H., Vivekananda, J., Ye, G. N., Tewari, K. K., and Sanford, J. C. (1990) Transient foreign gene expression in chloroplast of cultured tobacco cells following biolistic delivery of chloroplast vectors. *Proc. Natl. Acad. Sci. USA* **87**: 88-92.

91. Daniell, H., Krishnan, M., and McFadden, B. A. (1991) Expression of β -glucuronidase gene in different cellular compartments following biolistic delivery of foreign DNA into wheat leaves and calli. *Plant Cell Rep.* **9**:615-619.
92. FREY, J. (1999) Genetic flexibility of plant chloroplasts. *Nature* **398**:115-116.

IDENTIFICATION OF STRAWBERRY FLAVOUR RELATED GENES BY THE USE OF DNA MICROARRAYS

Asaph Aharoni and Ann P. O'Connell

Plant Research International, Business Unit Cell Cybernetics, P. O. Box 16, 6700 AA, Wageningen, The Netherlands.

1. INTRODUCTION

RNA-based assessment of gene expression provides essential clues to the elucidation of gene function. The current availability of complete genome sequences and of large sets of Expressed Sequence Tags (ESTs) from numerous organisms has triggered the development of efficient and accurate methods for large-scale and genome-wide analyses of genetic variation and gene expression patterns. As a result, several novel methods either sequence-based (Brenner *et al.*, 2000; Velculescu *et al.*, 1995), fragment-based (Bachem *et al.*, 1996; Shimkets *et al.*, 1999) or hybridisation-based such as macro- and microarrays (Desprez *et al.*, 1998; Lockhart *et al.*, 1996; Schena *et al.*, 1995) are currently available (see also Breyene and Zabeau, 2001).

Microarray technology is a hybridisation-based method combining miniaturisation and the application of fluorescent dyes for labelling. The latter facilitates the combination of two differently labelled samples in a single hybridisation experiment and thus the use of competitive hybridisation to reduce experimental error. In this way relative expression levels of large numbers of genes can be determined simultaneously with a

high degree of sensitivity. Today, two fundamentally different microarray-based technologies are available, both capable of large-scale expression analyses.

A photolithographic method for high-density spatial synthesis of oligonucleotides was introduced by Fodor and colleagues. With this method arrays can be produced containing up to a few hundred thousand distinct elements (Fodor *et al.*, 1991). As oligonucleotide arrays allow highly sensitive detection of DNA mismatches, they are well suited for DNA variation analysis as well. Manufacturing such arrays requires, however, prior sequence knowledge as well as complicated design and production methodologies (Lipshutz *et al.*, 1999). An alternative method, in which pre-synthesised nucleic acids are mechanically deposited onto a solid surface, allows a more flexible design for the fabrication of microarrays (Duggan *et al.*, 1999). In most cases PCR-amplified cDNA clones are used and the resulting arrays are referred to as cDNA microarrays. This technology however, can also be used to manufacture oligonucleotide arrays.

The basic microarray assay used with both types of array is similar and based on the specific hybridisation of a labelled sample (target) to the immobilised nucleic acids (probe) on the array. As a result, the complex mixture of nucleic acids isolated from the biological sample under study is spatially separated into its constitutive components, the specific mRNAs. The physical separation on the array enables the individual quantification of many specific mRNAs in a single hybridisation experiment. Furthermore, the independent detection of fluorescent signals at specific wavelengths allows simultaneous analysis of multiple dyes and thus mixed samples. Once data is collected and normalised, expression ratios are obtained for each individual gene, representing relative expression levels for the samples investigated. Ultimately, biological meaning is inferred from data analyses of the comparison between samples and genes across one or multiple experiments and the combination with related biological knowledge.

Expression profiling using microarrays is currently being performed for numerous organisms, including several plant species, using an assortment of biological samples. The scale of these experiments range from a few hundred genes to genome-wide coverage (e.g. *Saccharomyces cerevisiae*, Lashkari *et al.*, 1997; *Drosophila melanogaster*, Zou *et al.*, 2000; *Caenorhabditis elegans*, Jiang *et al.*, 2001). Next to its use for gene expression studies, microarrays are presently being widely applied to DNA variation analyses (Lander, 1999). Variation in DNA sequence underlies much of the phenotypic differences that can be observed, not

only between, but also within species and populations. Locating and identifying these genotypic differences allows linkage of genotypic and phenotypic variation. Point mutations, commonly referred to as Single Nucleotide Polymorphisms (SNPs), are the most frequent type of variation in genomes. Microarrays and in particular oligonucleotide arrays, may be used for large-scale SNP detection and discovery. In this report, we will focus on the use of microarrays produced by deposition of pre-synthesised nucleic acids (referred to as “cDNA microarrays”) and their application for gene expression monitoring. Photolithographic microarrays (referred to as “oligonucleotide microarrays”) and the analysis of DNA variation will be mentioned where appropriate.

In plants, reports on the use of cDNA microarrays to follow gene expression profiles of hundreds and thousands of genes have been growing exponentially. To date, most studies describing the use of microarrays have been performed with the model plant *Arabidopsis thaliana* (e.g. Ruan *et al.*, 1998; Wang *et al.*, 2000; Girke *et al.*, 2000; Harmer *et al.*, 2000; Maleck *et al.*, 2000; Reymond *et al.*, 2000; Zhu *et al.*, 2000; Schenk *et al.*, 2000; Schaffer *et al.*, 2001). The flexible nature of the fabrication and hybridisation methods of cDNA microarrays allows the application of the technology also to non-model species for the study of complex and scarcely investigated biological processes, such as fruit flavour and aroma biosynthesis (Aharoni *et al.*, 2000).

In the first part of this chapter we will describe the basic principles and current global applications of the microarray approach. The second part of the chapter will demonstrate the power of microarrays to link gene to function, and their use for studying pathways central to strawberry fruit flavour development. Volatile esters are important compounds providing fruity odors. Here we will describe the cloning of the *SAAT* gene encoding an ester-forming enzyme in the ripe strawberry fruit. The identification of other candidate flavour related genes through the association of gene expression patterns with existing biochemical knowledge will also be demonstrated.

2. DNA MICROARRAY TECHNOLOGY

2.1 Principle of the Method

Like other hybridisation-based analysis methods in molecular biology, the specificity of microarray technology relies on the selective and differential

hybridisation of nucleic acids. Earlier methods, such as DNA and RNA gel blot analysis, use a unique labelled nucleic acid molecule in solution. This so-called probe is hybridised to the complex mixture under study, such as a total RNA sample, that has been attached to a solid support. Information obtained from such experiments relates to the abundance of one single poly-nucleotide of interest. Array-based methods such as oligonucleotide arrays and cDNA arrays use the reverse strategy where complex mixtures of labelled poly-nucleotides (such as cDNA derived from mRNA) are hybridised with large numbers of individual elements (e.g. unique PCR-products in cDNA microarrays), attached to a solid surface. In this way information on the abundance of many poly-nucleotide species is gained in parallel.

Labelling with fluorescent dyes possessing different excitation and emission characteristics allows the simultaneous hybridisation of two samples on a single array. The strength of fluorescence emission at the two wavelengths represents the amount of a specific poly-nucleotide from each sample bound to the array. In this manner a single experiment provides quantitative hybridisation data for hundreds to thousands of probes.

For expression studies using cDNA microarrays the approach of combining two differently labelled samples (reference and test sample) is common practice. For each gene the corresponding amount of signal in both samples can then be quantified in parallel and expression ratios obtained. A schematic illustration of the cDNA microarray principle is shown in Figure 1.

2.2 cDNA Microarray Production

2.2.1 Preparation of Probes for cDNA Microarrays

The first step of cDNA microarray fabrication requires the selection of probes to be used. These can be PCR products resulting from the direct amplification of genomic DNA (by the use of gene specific primers) or amplified inserts from cDNA libraries (for instance ESTs) or any other library of interest. For example, the first microarray allowing genome-wide expression monitoring was generated by amplifying genomic DNA with specific primer pairs designed for 6200 open reading frames (ORFs) of yeast (DeRisi *et al.*, 1997). Prior to spotting, probes are often purified from unwanted PCR components and concentrated by precipitation or gel-filtration. Instead of PCR products (ranging in size from approximately

0.2-2.5 kb), large synthetic oligonucleotides (50 to 80 base pairs) can also be spotted for the purpose of gene expression studies (Mir and Southern, 1999; Kane *et al.*, 2000).

As the production of whole-genome arrays is still very expensive and laborious, a subset of genes can alternatively be used to make a dedicated array. By selecting the appropriate tissue, developmental stage or treatment as source material, a cDNA library enriched for genes involved in the process under study can be obtained and used to select clones. Another approach designed to enrich for targeted clones is to perform a pre-selection, identifying differentially expressed genes, such as suppression subtractive hybridisation (SSH; Diatchenko *et al.*, 1996; Yang *et al.*, 1999) or representational difference analysis (RDA; Welford *et al.*, 1998).

2.2.2 Printing cDNA Microarrays

Two different approaches both involving printing-type technologies are currently utilised in the fabrication of cDNA microarrays: contact printing (various methods for mechanical deposition) and non-contact printing (liquid delivery).

In contact printing, an array of either solid or split pins are dipped into the DNA solution for loading. A micro-droplet is subsequently deposited upon direct contact with the solid surface of the array (Figure 2A and 2B). The method uses a motion control system that spots or prints a precise sample of each probe onto multiple surfaces (often 50 to 100 microscope slides) in a serial operation. Depending on the application, contact printing usually produces sub-nanolitre droplets at a pitch of 100 to 250 μm . An example of a split spotting pin is shown in Figure 2C. Mechanical spotting as described above is, at present, the most common way for the fabrication of cDNA microarrays.

Non-contact printing involves the controlled ejection of small (nano to picolitres) volumes of DNA solution from a dispenser onto the surface from a defined distance. In contrast to contact printing, this method allows flexibility in printing volume. The most common type of non-contact dispensing uses various types of ink-jet technology (e.g. thermal, solenoid, piezoelectric) for droplet generation and delivery (Okamoto *et al.*, 2000). As for contact printing, both oligonucleotides and cDNAs can be handled. cDNA microarrays produced by ink jet printing may contain thousands of array elements.

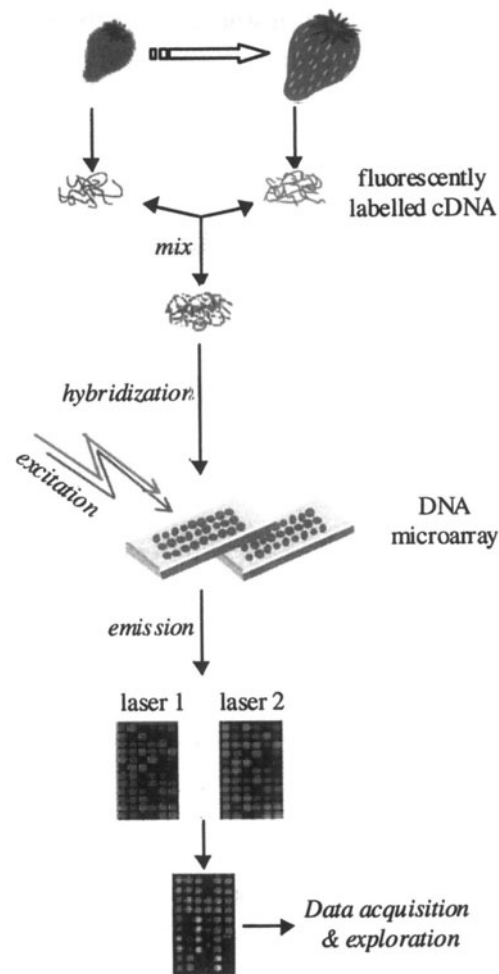


Figure 1. Scheme of a typical cDNA microarray assay for gene expression analysis. In this example mRNA levels are compared between the green and red stages of fruit development. First, mRNA is isolated from each tissue and reverse transcribed in the presence of different fluorescent dyes resulting in labelled cDNA. Next, the 2 cDNA populations are mixed and hybridised to a cDNA microarray. Each array element contains DNA representing a different gene. The specific cDNAs from both populations, representing individual transcripts, will hybridise specifically with the probe on the corresponding array element. After hybridisation, the microarray is scanned with a confocal laser device for fluorescence emission at two wavelengths following independent excitation of the two dyes. The relative abundance of mRNA from each gene in green vs. red fruit is reflected by the ratio green/red as measured by the fluorescence emitted from the corresponding array element. Image analysis software is used to determine fluorescence intensities that allow the quantitative comparison between the two stages of fruit development for all genes on the array.

cDNA microarrays are often fabricated on glass surfaces such as microscope slides. In order to enhance the adhesion of probes, to lower

the background and to restrict spreading of the droplets, the slides are pre-coated with e.g. poly-lysine or amino silanes. After spotting, the DNA is immobilised (either by UV cross-linking or baking), the unused surface is blocked (by succinic anhydride or sodium borate) and, as a final step, the DNA on the slide is denatured (by heat or alkali treatment). Processed slides may be stored dry for several months prior to hybridisation.

2.3 Labelling and Hybridisation

In a cDNA microarray experiment, samples under study are typically RNA preparations from two or more biological sources. The fluorescent labelling of mRNA is commonly performed by first-strand cDNA synthesis in the presence of modified nucleotides using oligo-dT as a primer. Including fluorescently labelled nucleotides during the reverse transcription reaction results in the direct synthesis of labelled cDNA.

The fluorescent labels Cyanine-3 and Cyanine-5 are frequently paired, as they possess relatively high incorporation efficiencies with reverse transcriptase, good photostability and yield, and light absorption and emission at distinct and separable wavelengths. Both total RNA and mRNA may be used as the starting material for labelling, although the use of the latter provides the best hybridisation results. To obtain the desired fluorescent signal, 10 μg to 50 μg of total RNA or 0.5 μg to 2.5 μg of mRNA is used, per sample, per array. Improvements in labelling schemes such as by target amplification allows reduction of the amount of RNA required (0.1 μg total RNA) and the use of a minimum amount of tissue. This will facilitate studying gene expression samples derived from just a few cell layers (Hertzberg *et al.*, 2001). A mixture of equal amounts of both labelled samples (between 5 μl to 50 μl total volume) are hybridised to the array usually under a coverslip. The slide is then placed in a specially designed reaction chamber to avoid evaporation. The hybridisation conditions, such as ionic strength, temperature and target concentration, depend on the application. Hybridisation temperatures of 42 °C (when hybridising in 50% formamide) to 70 °C (when using SSC-based buffers) for several hours to overnight are routinely employed.

2.4 Acquisition of Microarray Expression Data

Once the fluorescent sample is hybridised to a cDNA microarray, unbound material is washed away and the sample hybridised to each element is visualised by fluorescence detection. Both confocal scanning devices and CCD cameras are being used for this purpose. Fluorescence

emission from the microarray is converted into a digital output for each dye and is stored as separate image files. Next, image analysis software is used for quantification of individual array elements. A grid is superimposed over the image and the average (or median) pixel intensities for each element is calculated for both dyes (Figure 2D). Background fluorescence is then subtracted from the raw data and the figures are normalised to correct for channel specific effects, such as differences in quantum yield of the dyes and unequal labelling efficiencies of the samples. Normalisation also corrects for any unwanted differences in the amount of sample used.

2.5 Data Exploration

The overwhelming amount of data generated by microarray experiments poses in itself problems to the average plant molecular biologist not yet acquainted with handling large data sets. These problems range from important but trivial subjects, such as clone tracking (i.e. relocating individual clones on the array) and data storage, to more sophisticated topics such as the visualisation and analyses of multiple-sample experiments and general data mining.

Several standard statistical techniques are currently being used to help interpreting microarray data, including hierarchical clustering, principal component analysis (PCA) and self-organising maps (SOM). These techniques are all focused on grouping together genes (or samples) which show similar behaviour. This type of analysis, using large data sets, can provide novel perspectives on cellular regulatory mechanisms and can associate expression of unknown genes with a putative function. Hierarchical clustering of gene expression data in combination with false colour coding of the expression levels has become a popular way of data analysis and presentation (Eisen *et al.*, 1998). With this technique genes are grouped in clusters based on the similarity between their expression profiles. In a bottom up approach genes are joined to form nodes, which in turn are then further joined. Joining proceeds until all genes are combined in a single hierarchical tree.

2.6 The Application of DNA Microarrays in Gene Expression Analysis

As for other biological disciplines, plant microarrays will prove an indispensable tool for research in the field of molecular plant science. The systematic, non-biased, accurate and large-scale acquisition of data using

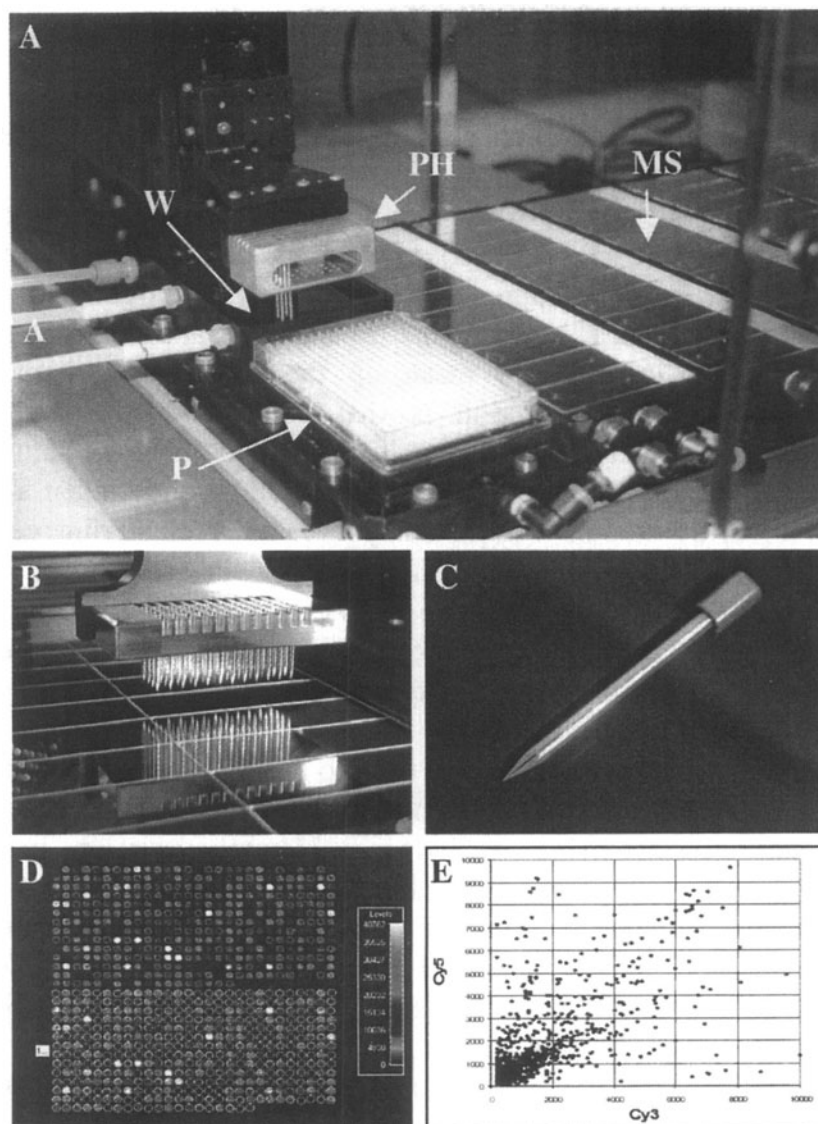


Figure 2. Fabrication of cDNA microarrays by contact printing. (A) Motion control system printing a sample of each probe onto multiple microscope slides (MS) in a serial operation [PixSys 7500 arrayer (Cartesian Technologies) provided with ChipMaker microspotting print head (PH) (TeleChem International, Inc)]. P- 384 well plate, W- wash station. (B) Print head used for contact printing containing an array of 48 split pins (ChipMaker). (C) Detail of a split spotting pin. (D) Extracting the signal intensity of each individual array element. Image analysis software is used to superimpose a grid onto the image as shown in the lower half of the array. (E) Graphical representation of a typical raw data set obtained from a single microarray experiment. Each spot in the graph corresponds to an individual array element. Figures (B) and (C) are courtesy of T. Martinsky, TeleChem International, Inc.

microarray technology enables new experimental approaches for plant molecular biologists. Microarray technology already provides, even at this early stage of its application in plant research, a global overview of biological mechanisms which until just recently were investigated in a “gene by gene” manner.

Presently gene expression monitoring is the most widespread application of microarrays. Microarray assays may be directly integrated into functional genomic approaches aimed both at assigning function to identified genes, and at studying the organisation and control of genetic pathways acting together to make up the functional organism. The rationale behind this approach is that genes showing similarity in expression pattern may be functionally related and under the same genetic control mechanism. Therefore, a common strategy undertaken already in early microarray studies was to analyse data by clustering genes into groups, based on their expression profiles as scored in multiple experiments (Brown and Botstein, 1999). In most cases, gene clusters comprise both known and unknown genes, allowing researchers to associate putative functions to the unknown genes by employing the concept of “guilt by association”.

Essentially, microarrays may be used to analyse any kind of variability in gene expression between given samples. These differences can be either naturally occurring or induced. Natural variation may occur between different plant cultivars, tissues, developmental stages, environmental conditions or during circadian rhythm. Induced changes in gene expression may arise from experimental exposure to different environmental conditions or result from mutagenesis. Microarray experiments in which the response to drought and cold stresses (Seki *et al.*, 2001), mechanical wounding and insect feeding (Reymond *et al.*, 2000), herbivory (Arimura *et al.*, 2000) and nitrate treatments (Wang *et al.*, 2000) was investigated, have already demonstrated the capability of microarray-assisted expression studies to identify novel response genes including those encoding regulatory factors.

In expression monitoring, a single hybridisation experiment can provide quantitative results for 18,000 (Alizadeh *et al.*, 2000) genes simultaneously. The high density and miniaturisation make genome-wide expression studies feasible, by using either cDNA or oligonucleotide arrays. Genome-wide expression profiling at the transcript level is one of the most exciting tools to study the cell and its integrative processes. The ability to monitor simultaneously the expression of a large set of genes is one of the main spin-offs of genome sequencing efforts. Current reports on genome-wide expression analysis in plants also describe the use of

microarrays (either oligonucleotide or cDNA) and already cover approximately one-third of the *Arabidopsis* genome (Wisman and Ohlrogge, 2000; Zhu and Wang, 2000). During the coming years results from hundreds of microarray experiments will be collected in special gene expression databases [e.g. Stanford Microarray Database (<http://genomewww4.stanford.edu/MicroArray/SMD>)]. The flexible nature of the arrays makes them applicable for the exploration of more complex genomes (e.g. non-model species such as strawberry), in order to answer biological questions that simply cannot be answered by analyzing model species such as *Arabidopsis*. For example, one can now study the genetic controls governing the biosynthesis of commercially important plant secondary metabolites, such as flavour and aroma compounds.

3. STRAWBERRY FRUIT RIPENING AND FLAVOUR

3.1 The Biosynthesis of Esters and Other Flavour Compounds in Strawberry During Fruit Ripening

Fruit are an important part of our diet mainly as a source of energy, vitamins, and minerals in addition to their aesthetic qualities like color, flavour, and taste. Unlike fruit botanically defined as arising from the expansion of the ovary, strawberry is actually the swollen base of the flower (receptacle) with one seeded fruit (termed achenes) located on the outer surface. Strawberry is classified as a non-climacteric fruit since it does not show the typical increase in ethylene levels and respiration preceding the ripening process of for example tomato and melon, which are classified as climacteric fruits (Manning, 1994). Early on in strawberry research, it was demonstrated that the decline in auxin levels supplied from the achenes to the receptacle tissue during fruit development was associated with the onset of strawberry fruit ripening (Given *et al.*, 1988). Later, strong evidence was provided that auxin triggered ripening by inducing expression of ripening related genes (Manning, 1994; Manning, 1998; Harpster *et al.*, 1998; Moyano *et al.*, 1998). Strawberry displays the same dramatic changes during ripening as other fruit. Fruit ripening is characterised by a rise in soluble solid content in the receptacle, the production of natural aroma and flavour compounds and alterations of fruit shape, size, texture and pigmentation (Woodward, 1972).

In general fruit flavour compounds develop during ripening when the metabolism of the fruit changes towards catabolism. Analogous to other fruit, a complex mixture of hundreds of compounds (Zabetakis and Holden, 1997) determines strawberry flavour and aroma (Figure 3). More than 300 compounds have been identified that can contribute to the unique aroma/flavour of ripe strawberry fruit. The components identified may be grouped into several chemical classes, which include acids, aldehydes, ketones, alcohols, esters, and lactones. Other contributing groups are sulphur compounds, acetals, furans, terpenes, phenols and epoxides. Members of these groups, often present at low levels, may have a significant impact on the overall flavour of strawberry. The compound 2,5-dimethyl-4-hydroxy-3(2H)-furanone (furanol) and its methyl ether, mesifuranol (2,5-dimethyl-4-methoxy-3(2H)-furanone) were the only compounds identified in diluted solutions, as exhibiting the typical flavour/aroma associated with strawberry (Roscher *et al.*, 1997). The biochemical pathway leading to furaneol and its methyl ether is unknown although there is strong evidence that sugars (most likely fructose-1,6-bisphosphate) supply the precursors required for furaneol biosynthesis (Schwab, 1998).

3.2 Biosynthesis of Volatile Esters in Strawberry

Volatile esters are quantitatively and qualitatively the most important compounds providing fruity odors, and in strawberry alone more than a hundred different esters have been detected (Maarse, 1991). A biosynthetic pathway of volatile esters in fruit has been proposed previously (Olias *et al.*, 1995) (Figure 4). Figure 4 shows that esters in strawberry are derived from various precursors from primary metabolism such as fatty acids and amino acids. The oxidative degradation of membrane lipids to linolenic and linoleic acid catalysed by lipoxygenase is the main source of precursors for the generation of alcohols, aldehydes, acids and esters found in fruit (Perez *et al.*, 1999). The last enzyme in the β -oxidation of fatty acids is thiolase that catalyses the thiolytic cleavage involving another molecule of CoA. The product of this reaction is acetyl-CoA and acyl-CoA derivatives containing two carbon atoms less than the original acyl-CoA molecule that underwent oxidation. The acyl-CoA formed in the cleavage reaction can be utilized as one of the two co-substrates for the final step of ester formation in fruit (Bojorquez *et al.*, 1995).

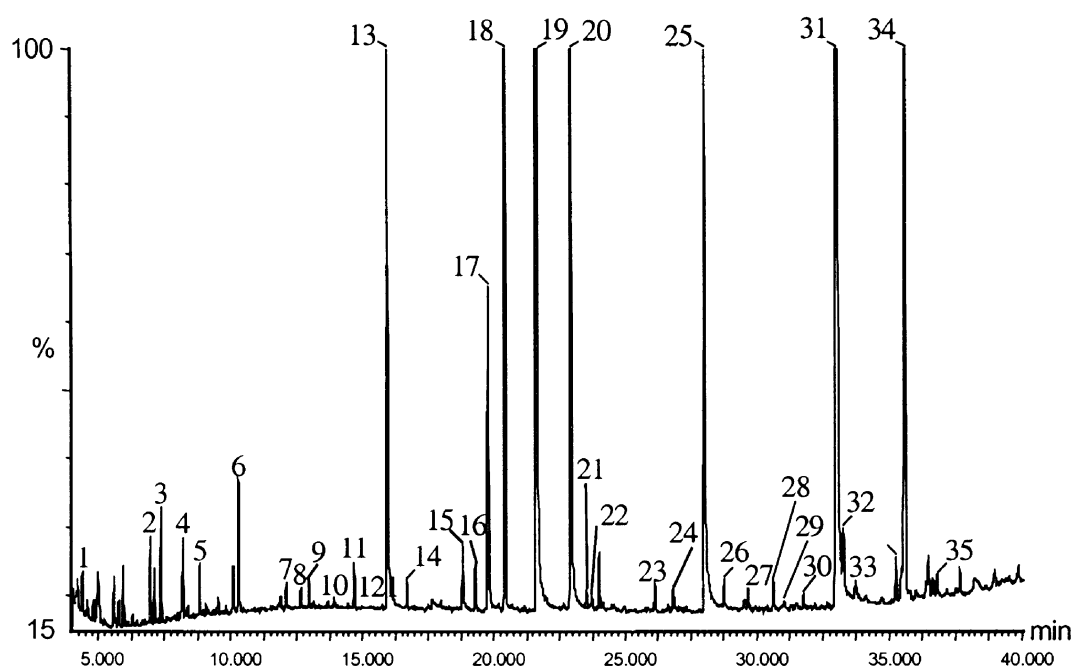


Figure 3. Volatile constituents of ripe strawberry fruit (cv. Elsanta) detected by GC-MS after XAD solid phase extraction. Peak numbers correspond to: 1- butyl acetate; 2- 4 methyl-2-butenal; 3- methyl hexanoate; 4- *E*-2-hexanal; 5- ethyl hexanoate; 6- 3-hydroxy-2-butanone; 7- *E*-2-hexenyl acetate; 8- *E*-rose oxide; 9- hexanol; 10- *Z*-3-hexenol; 11- *E*-2-hexenol; 12- *E*-linalooloxid, furanoid; 13- acetic acid; 14- *Z*-linalooloxid, furanoid; 15- propanoic acid; 16- linalool; 17- 2-methyl propionic acid; 18- methoxyfuraneol; 19- butanoic acid; 20- 2-methyl butanoic acid; 21- gamma-hexalactone; 22- 4-ethyl benzaldehyde; 23- delta-hexalactone; 24- 4-methyl pentanoic acid; 25- hexanoic acid; 26- benzylalcohol; 27- gamma octalactone; 28- benzothiazole; 29- delta-octalactone; 30- furaneol acetate; 31- furaneol; 32- *E*-nerolidol; 33- octanoic acid; 34- gamma-decalactone; 35- delta-decalactone. (The image is courtesy of W. Schwab).

In addition to fatty acids, the transamination of amino acids provides the precursors for volatile aroma compounds such as aldehydes, acids, alcohols, esters and thiols (Perez *et al.*, 1992). In strawberry, alanine is proposed to be the main free amino acid metabolized to flavour compounds (Perez *et al.*, 1992). This was based mainly on the dramatic decrease in its content just before the formation of volatile aroma compounds had commenced in the fruit. In addition, feeding of alanine to strawberry cultures resulted in the formation of several esters such as methyl and ethyl hexanoate, which are important constituents of strawberry (Perez *et al.*, 1992). The transamination of amino acids is catalyzed by aminotransferases. Ketoacids produced by transamination can be enzymatically degraded to the corresponding aldehydes. The

enzyme that could catalyze this reaction is pyruvate decarboxylase. Alcohol dehydrogenase enzymes have been implicated in the interconversion of the aldehyde and alcohol forms of flavour volatiles (Scharpf and Chandan, 1989). Esterification is the result of transacylation from acyl-CoA to an alcohol (Perez *et al.*, 1996). The enzyme catalyzing this reaction is called an alcohol acyltransferase (AAT).

3.3 Alcohol Acyltransferase and Ester Formation

Alcohol acyltransferases have been identified in both yeast and fungi (Fujii *et al.*, 1996; Fujii *et al.*, 1994; Yamakawa *et al.*, 1978). The yeast alcohol acyltransferases showed high affinity to acetyl-CoA and the alcohols ethanol and isoamyl alcohol. The influence of esters (isoamyl- and ethylacetate) on beer flavour renders AAT one of the most important enzymes in the fermentation process performed by microorganisms (Yoshioka and Hashimoto, 1984). In plants, ester formation has been studied in both fruit and flowers. In *Clarkia breweri* the ester benzylacetate is an important constituent of the flower scent. The purification of the acetyl-CoA:benzylalcohol acetyl-transferase (BEAT) protein from flowers of *C. breweri* and the isolation of the gene encoding it has been reported (Dudareva *et al.*, 1998). BEAT has a high affinity for aromatic alcohols, such as benzyl alcohol and cinnamyl alcohol.

In fruits such as melon, strawberry and banana, AAT proteins have been investigated using crude fruit extracts (Harada *et al.*, 1985; Ueda *et al.*, 1992; Perez *et al.*, 1993; Olias *et al.*, 1995; Perez *et al.*, 1996). Analysis of the substrate specificity of the enzymes from the various sources, revealed differences in their affinity to acyl-CoAs and alcohols. Maximum activity for the strawberry AAT was obtained using acetyl-CoA and hexanol as substrates (Olias *et al.*, 1995).

4. NOVEL INSIGHT INTO STRAWBERRY FLAVOUR FORMATION USING CDNA MICROARRAYS

4.1 Identification of the SAAT Gene

As previously discussed, several metabolic pathways are involved in the biosynthesis of flavour and aroma compounds in fruit. Some of these

pathways have barely been investigated even at the biochemical level and this makes the identification of flavour related genes more complicated.

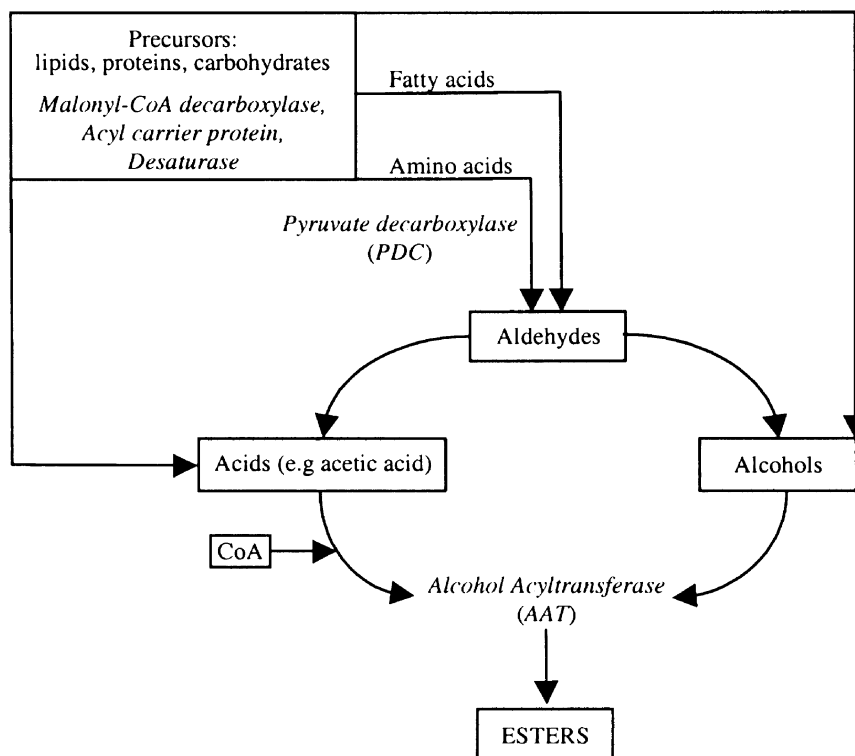


Figure 4. Scheme showing probable metabolic routes leading to the formation of volatile esters in fruit. Genes identified by our microarray study showing similar expression profile during strawberry ripening are depicted (in *Italic*). Adapted and modified from Olías et al., (1995).

During fruit development and ripening, both physical as well as morphological changes are often the result of changes in protein levels and activities, which may reflect shifts in overall mRNA abundance. We anticipated that the dramatic changes related to flavour and aroma biosynthesis in fruit are in part also the result of changes in the abundance of certain mRNAs. Thus microarray assays were our method of choice to discover genes associated with flavour formation in strawberry, as they allow sensitive, quantitative, large-scale, rapid and parallel monitoring of gene expression levels. We hypothesised that genes central to flavour formation would be discovered by correlating the expression profile of known ripening related genes over several developmental stages with candidate genes deduced from biochemical studies on flavour formation. Pigmentation related genes serve as good molecular markers since colour

changes in the majority of fruit normally are accompanied by the accumulation of flavour and aroma components (Seymour et al. 1993).

An overview of our microarray approach is depicted in Figure 5. Initial steps involved performing detailed cytological analyses of developing strawberry fruit, in order to examine fruit cell structure and understand physiological and morphological events associated with development and maturation. In parallel with the cytological observations, we conducted biochemical analyses using gas chromatography coupled to mass spectrometry (GC-MS), to examine in detail the composition of flavour/volatile components, and the temporal and spatial patterns of their production and emission. Subsequently, a cDNA library was constructed using mRNA derived from red ripe fruit of the domesticated octaploid variety *Elsanta* (*Fragaria* x *ananassa*). Mass excision of 1100 cDNAs from the cDNA library was performed and sequences obtained by a single sequence reaction (random sequencing, ESTs) were examined for homology to sequences present in the public databases. In the next step, the 1100 cDNAs plus an additional 601 un-sequenced cDNAs were spotted on glass microscope slides.

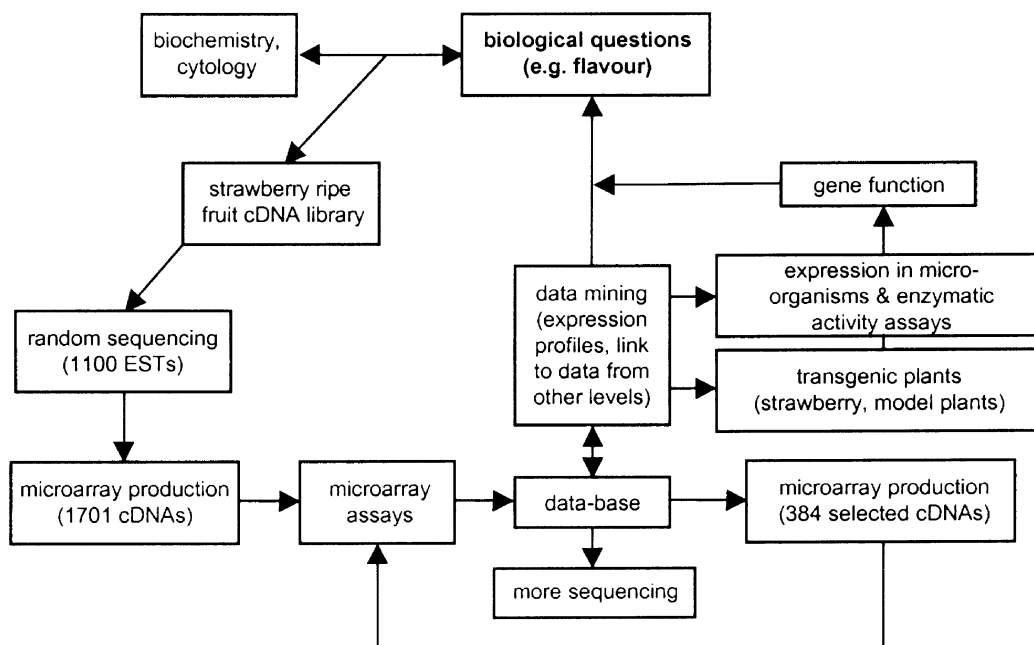


Figure 5. The microarray approach used to identify flavour/aroma related genes in strawberry.

The first three microarray experiments compared strawberry fruit developmental stages a) green with red b) white with red and c) turning with red. In each experiment two mRNA populations fluorescently labelled with cyanine 3 (Cy3) or cyanine 5 (Cy5) were hybridized to the microarray simultaneously. After hybridization, the fluorescence of each individual array element was recorded at the specific wavelengths of each fluorescent dye and calculation of expression ratios was performed for each cDNA (Aharoni *et al.*, 2000). The expression profile of an unknown gene (*SAAT*) closely matched the expression profile of known ripening regulated genes (e.g. pigmentation and cell-wall related genes). Quantitative microarray expression analysis revealed that *SAAT* had a 16-fold higher expression level during the red stage of fruit development than in the green stage. The full-length *SAAT* cDNA clone is 1618 bp, encoding a polypeptide of 452 amino acid residues with a predicted molecular mass of 50.7 kDa. The presence of several consensus sequences in the *SAAT* protein assigned it to a super-family of multifunctional acyltransferases responsible for coenzyme A-dependent acyl transfer (St-Pierre *et al.*, 1998). As the cDNA library (from which the probes arrayed were selected) was prepared from whole fruit containing a mixture of achene and receptacle tissue, we performed a fourth experiment to compare gene expression between the ripening stage achene and receptacle tissues. This experiment showed that *SAAT* expression was associated with the receptacle tissue. Detailed RNA gel blot analysis confirmed the microarray result and showed that *SAAT* gene expression was receptacle-specific commencing during the white stage of fruit ripening, preceding detectable volatile ester formation in strawberry, and reaching maximal levels between turning and red stages (Figure 6).

4.2 *SAAT* Encodes the Ester-Forming Enzyme from Strawberry Fruit

Biochemical evidence that the *SAAT* gene is involved in the formation of fruity esters was obtained by the characterisation of the recombinant protein expressed in *Escherichia coli* (Figure 7). The *SAAT* enzyme showed maximum activity with aliphatic medium chain alcohols such as hexanol and octanol, whose corresponding esters are major components of strawberry volatiles. It was capable of utilising short and medium chain, branched and aromatic acyl-CoA molecules as co-substrate. The *SAAT* enzyme was also capable of using the 10-carbon decanoyl-CoA as a co-substrate. This ability may account for the presence of this type of long chain esters in strawberry (Honkanen and Hirvi 1990). Clear differences

in enzyme activity were detected between the four isomers of hexenol tested: *trans*-2-hexenol was a better substrate than *cis*-2-hexenol, while *cis*-3-hexenol was a better substrate than *trans*-3-hexenol. The effect of the position of the hydroxy moiety of the alcohol on SAAT activity varied: 1-propanol was a better substrate than 2-propanol, however, 2-butanol was a better substrate than 1-butanol. SAAT also accepted the branched primary alcohol isoamyl alcohol. Activity was also detected with aromatic (benzyl- and phenylethyl-) and cyclic (furfuryl-) alcohols, though activities were much lower than with 1-octanol (4 to 10%). In contrast, no activity could be detected with the terpene alcohol linalool. This is in agreement with the absence of linalool esters in strawberry volatile profiles, including the *Elsanta* cultivar that does contain linalool (Honkanen and Hirvi 1990).

4.3 Other Candidate Genes Associated with Flavour Formation in Strawberry

The identification of *SAAT* provided us with a valuable expression profile to assist in the identification of other putative flavour associated genes in strawberry. As described previously, fatty acids serve as the initial precursors for several groups of flavour and aroma compounds present in many fruits including strawberry. Aliphatic C-6 compounds (aldehydes, alcohols, acids and esters) contributing to strawberry fruit flavour are formed from unsaturated aliphatic C-18 fatty acids, linoleic (C_{18:2}) and linolenic acid (C_{18:3}), through the lipoxygenase/hydroperoxide lyase pathway (Croteau and Karp, 1994). Plant fatty acid unsaturation begins with the conversion of 16:0 acyl carrier protein (ACP) and 18:0-ACP into 16:1-ACP and 18:1-ACP, respectively, by $\Delta 9$ stearoyl-ACP desaturase (Wang *et al.*, 1996). From our microarray developmental study we observed high expression levels for genes encoding a strawberry $\Delta 9$ desaturase, an acyl carrier protein (ACP) and a gene putatively involved in fatty acid biosynthesis (malonyl-CoA decarboxylase) during the red stage of fruit development. The expression pattern of a putative strawberry pyruvate decarboxylase (*PDC*) gene was also investigated, because it may play a role in providing precursors for the formation of ethyl esters. The *PDC*, $\Delta 9$ stearoyl-ACP desaturase, malonyl-CoA decarboxylase and ACP expression profile correlated well with the *SAAT* expression profile during fruit development. Interestingly, the activities of two other intermediate enzymes in the metabolism of fatty acids, lipoxygenase and hydroperoxide lyase, were previously reported to increase steadily from

the white to the red stage of strawberry fruit development (Perez *et al.*, 1999).

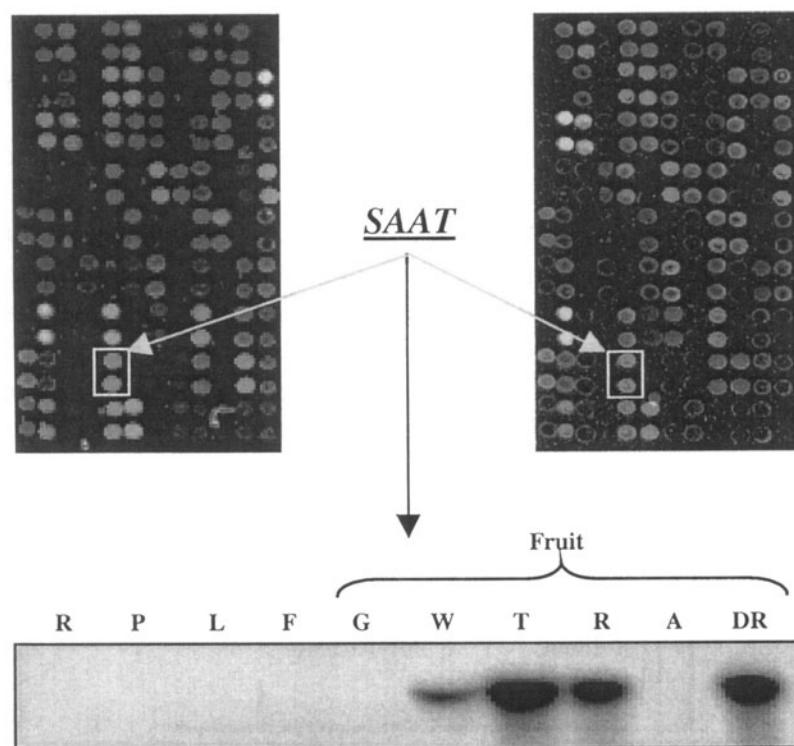


Figure 6. *SAAT* gene expression detected by cDNA microarray assays and RNA gel blot analysis. The microarray images are a two-color overlay obtained with green stage target and red stage target cohybridized with a single microarray (left subarray) and receptacle target and achene target cohybridized with another microarray (right subarray). In the superimposed images the green stage and achene targets are represented as a green signal, and the red stage and receptacle targets as a red signal. Signal intensities provide an estimate of expression levels and green spot colors correspond to higher transcript levels in the green stage or achenes and red spot colors to the red stage or receptacle targets. Genes with no significant difference in expression between the two stages of development or the two tissues show an intermediate yellow or brown color. The *SAAT* cDNA appeared as an intense red signal in both (boxed) and is marked by an arrow (each cDNA was arrayed in duplicate). The physical size of the subarrays is approximately 4 x 2.5 mm. Below RNA gel blot analysis of the expression of *SAAT* in different tissues of strawberry. R, root; P, petiole; L, leaf; F, flower; G, green fruit; W, white fruit; T, turning fruit; R, red fruit; A, achenes; DR, dark red fruit.

As mentioned previously, the decline in auxin levels in the receptacle tissue during fruit development (approximately at the end of the green stage) triggers ripening by inducing the expression of ripening related genes (Manning, 1998). By treating green strawberry fruit *in planta* with exogenous auxin (NAA at a concentration of 0.5 mM in lanolin paste), one could suppress the transcription of ripening related genes. In this way, auxin repressed ripening related genes would be those normally up-regulated in the receptacle during ripening. Using the second type microarray produced we compared gene expression in auxin treated and non-treated fruit.

Thirty-two ripening regulated cDNAs were repressed by the auxin treatment, of which 24 are depicted in Table 1. Fruit ripening processes that appeared to be auxin dependent included pathways related to pigmentation, stress/defence, cell wall metabolism, cell structure, fatty acid metabolism and flavour. The dramatic repression of *SAAT* in this experiment (6.5 fold) suggests that ester formation in strawberry is triggered and controlled by similar signalling pathways as for other typical ripening related processes such as pigmentation.

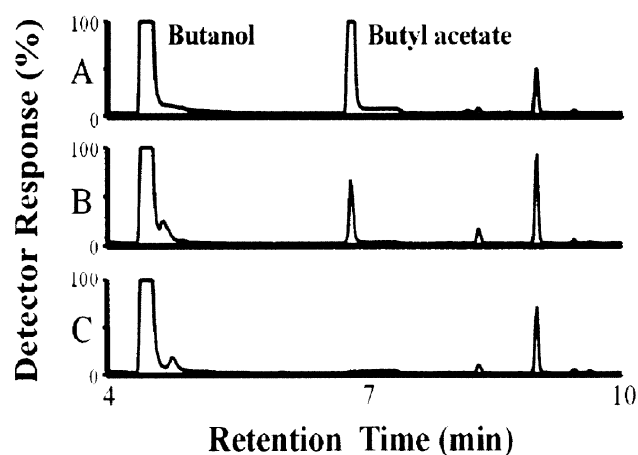


Figure 7. Verification of ester formation by the SAAT recombinant protein produced in *E.coli* using GC-MS. GC-MS (detector response, 100% = 2×10^6 total ion counts) of volatiles produced with the incubation conditions as described below. (A) Butanol and butyl acetate standards, (B) SAAT protein plus butanol and acetyl-CoA and (C) As given for (B); protein absent.

Interestingly, in the climacteric fruit melon, blocking ACC oxidase gene expression and thus ethylene biosynthesis resulted in a pronounced effect on their aroma profile mainly through a decrease in ester formation (Bauchot *et al.*, 1998). This indicates a similarity in the regulatory

processes related to volatile ester formation between climacteric (ethylene triggered) fruit and non-climacteric fruit such as strawberry (auxin triggered). The ACP (D117), malonyl-CoA decarboxylase (JB 19) and pyruvate decarboxylase (H51), of which the expression profiles correlated well with the *SAAT* expression profile during development and ripening, were also repressed (1.9, 2, 2 fold, respectively) by the auxin treatment. The $\Delta 9$ desaturase did not show any change in expression and might therefore be under the control of different signalling and regulatory pathways.

As for many other secondary metabolites, flavour/aroma components are often modified by methylation and glycosylation (Wintoch *et al.*, 1991). Furaneol, one of the main flavour compounds in strawberry, is very often detected in its methylated form known as mesifuraneol (Schwab, 1998). Methylation of furaneol might be performed by enzymes (O-methyltransferase) similar to the one encoded by cDNA F102, showing ripening regulated expression and repression by the auxin treatment (Table 1).

Another gene (JB136) that is encoding a farnesyl diphosphate synthase (FDP synthase), repressed by the auxin treatment (2.9, fold), might influence the biosynthesis of sesquiterpenoids, an additional group of flavour components detected in strawberry (Table 1). During strawberry ripening, the accumulation of the sesquiterpenoid nerolidol could be detected (Figure 3). The enzyme FDP synthase catalyses the reaction forming the C15 molecule FPP by condensation of a C5 molecule (isopentenyl diphosphate, IPP) to a C10 molecule (geranyl diphosphate, GPP). FPP may be further converted to nerolidol by the activity of a sesquiterpene synthase enzyme (Bouwmeester *et al.*, 1999).

The role for the above candidate genes in flavour and aroma biosynthesis in strawberry fruit remains to be verified. This can be achieved by introducing them into strawberry or model plants (antisense and/or overexpression approach), or by expression in microorganisms and enzymatic activity assays with the recombinant proteins.

5. CONCLUDING REMARKS

Despite the fact that DNA microarray technology is still in its infancy, many authors have now demonstrated its power by using it for a wide range of applications such as monitoring gene expression and mutation detection. The identification of the *SAAT* gene involved in fruit flavour biogenesis is a prime example of how microarrays can be used, to aid in

the dissection and exploration of metabolic pathways that lead to the formation of flavour constituents, and as a tool to link genes to function.

Table 1. Auxin repressed cDNAs as detected by cDNA microarray experiments.

PRI Clone	Auxin Repressed (fold) ^a	R/A ^b	Homolog Definition ^c	Putative Function
A104	5.9	8.4	Glutathione S-transferase	Anthocyanin sequestration
<i>CHS</i>	6.7	22.2	Chalcone synthase	Flavonoid pathway
C122	6.5	4.3	Profilin	Actin binding
C23	2.3	9.7	Ribosomal protein L13E	Unknown
D117	1.9	3.7	Acyl carrier protein	Fatty acid synthesis
E149	1.9	5.0	Quinone reductase-like	Reduction of quinones
E30	1.8	8.5	Pectate lyase	Cell wall degradation
E80	2.2	8.7	Endo-1,4-beta-glucanase	Cell wall degradation
F102	3.6	8.4	O-methyltransferase	Secondary metabolites methylation
F157	2.6	20.3	Chalcone-flavonone isomerase	Flavonoid pathway
F193	2.0	3.3	Cinnamyl alcohol dehydrogenase	Lignin biosynthesis
G175	2.1	3.5	Isoflavone reductase like	Phytoalexin synthesis
G84	7.2	3.7	Beta-tubulin	Cytoskeleton
H142	8.7	21.1	Dioxygenase	Unknown
H159	4.8	3.6	Ripening-induced protein	Unknown
H51	2.0	6.2	Pyruvate decarboxylase	Aldehyde formation
H61	3.3	16.0	Flavanone 3-hydroxylase	Flavonoid pathway
JB136	2.9	4.6	Farnesyl diphosphate synthase	Terpene metabolism
JB173	2.5	12.6	Chalcone reductase-like	Phytoalexin synthesis
JB19	2.0	4.6	Malonyl-CoA decarboxylase	Fatty acid synthesis
JB202	3.0	4.0	Cysteine proteinase	Protein breakdown
JB77	3.6	19.1	Anthocyanidin synthase	Flavonoid pathway
<i>SAAT</i>	6.5	16.3	Alcohol acyltransferase	Ester formation

^aExpression ratio detected in microarray experiment examining response to auxin (fold repression).

^bExpression ratio in the microarray experiment comparing expression in receptacle tissue (R) vs. achene tissue (A).

^cDefinition of the first BLAST X homolog. Comparison analysis of the sequences was conducted with the advanced basic local alignment search tool, BLAST server (Altschul *et al.*, 1990) and the National Center for Biotechnological Information (www.ncbi.nlm.nih.gov) non-redundant protein database.

In this chapter we also describe the identification of new candidate flavour related genes by a similar principle used to identify *SAAT*. Their role in strawberry flavour formation remains yet to be explored.

Since the isolation of the *SAAT* gene, we have made major strides in the identification of alcohol acyltransferases from a number of other fruits (e.g. banana, apple and melon). The identification of fruit alcohol acyltransferases provides a natural alternative route for biotechnology production of bio-flavours (novel flavours) and natural aroma chemicals in the form of volatile esters. It also provides the means to restore flavour characteristics of commercial fresh fruits lost through successive years of breeding and selection.

6. ACKNOWLEDGEMENTS

We would like to thank Harro Bouwmeester and Raffaella Greco for critically reading the manuscript.

7. REFERENCES

- Aharoni, A., Keizer, L.C.P., Bouwmeester, H.J., Sun, Z.K., Alvarez Huerta, M., Verhoeven, H.A., Blaas, J., van Houwelingen, A., De Vos, R.C.H., van der Voet, H., Jansen, R.C., Guis, M., Mol, J., Davis, R.W., Schena, M., van Tunen, A.J., and O'Connell A.P. 2000. Identification of the *SAAT* gene involved in strawberry flavour biogenesis by use of DNA microarrays. *Plant Cell* 12: 647-661.
- Alizadeh, A.A., Eisen, M.B., Davis, R.E., Ma, C., Lossos, I.S., Rosenwald, A., Boldrick, J.G., Sabet, H., Tran, T., Yu, X., Powell, J.I., Yang, L.M., Marti, G.E., Moore, T., Hudson, J., Lu, L.S., Lewis, D.B., Tibshirani, R., Sherlock, G., Chan, W.C., Greiner, T.C., Weisenburger, D.D., Armitage, J.O., Warnke, R., Levy, R., Wilson, W., Grever, M.R., Byrd, J.C., Botstein, D., Brown, P.O., and Staudt, L.M. 2000. Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature* 403: 503-511.
- Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. 1990. Basic local alignment search tool. *J. Mol. Biol.* 215: 403-410.
- Arimura, G., Tashiro, K., Kuhara, S., Nishioka, T., Ozawa, R., and Takabayashi, J. 2000. Gene responses in bean leaves induced by herbivory and by herbivore-induced volatiles. *Biochem. Biophys. Res. Commun.* 277: 305-310.
- Bachem, C.W.B., Van Der Hoeven, R.S., De Bruijn, S.M., Vreugdenhil, D., Zabeau, M., and Visser, R.G.F. 1996. Visualization of differential gene expression using a novel method of RNA fingerprinting based on AFLP: Analysis of gene expression during potato tuber development. *Plant J.* 9: 745-753.

- Bauchot, A.D., Mottram, D.S., Dodson, A.T., and John, P. 1998. Effect of aminocyclopropane-1-carboxylic acid oxidase antisense gene on the formation of volatile esters in *Cantaloupe Charentais* melon (cv. *Vedrandais*). *J. Agric. Food Chem.* 46: 4787-4792.
- Bojorquez G., and Gomez-Lim M.A. 1995. Peroxisomal thiolase mRNA is induced during mango fruit ripening. *Plant Mol. Biol.* 28: 811-820.
- Bouwmeester, H.J., Verstappen, F.W.A., Posthumus M.A., and Dicke, M. 1999. Spider mite-induced (3S)-(E)-nerolidol synthase activity in cucumber and lima bean. The first dedicated step in acyclic C11-homoterpene biosynthesis. *Plant Physiol.* 121: 173-180.
- Brenner, S., Johnson, M., Bridgham, J., Golda, G., Lloyd, D.H., Johnson, D., Luo, S.J., McCurdy, S., Foy, M., Ewan, M., Roth, R., George, D., Eletr, S., Albrecht, G., Vermaas, E., Williams, S.R., Moon, K., Burcham, T., Pallas, M., DuBridge, R.B., Kirchner, J., Fearon, K., Mao, J., and Corcoran, K. 2000. Gene expression analysis by massively parallel signature sequencing (MPSS) on microbead arrays. *Nature Biotech.* 18: 630-634.
- Breyne, P., and Zabeau, M. 2001. Genome-wide expression analysis of plant cell cycle modulated genes. *Curr. Opin. Plant Biol.* 4: 136-142.
- Brown, P.O., and Botstein, D. 1999. Exploring the new world of the genome with DNA microarrays. *Nature Genet.* 21 Suppl.: 33-37.
- Croteau R., and Karp F. 1994. Origin of Natural Odorants. In: Muller PM and Lamparsky D (eds.) *Perfumes : art, science, and technology*. Blackie Academic & Professional, an imprint of Chapman & Hall, Glasgow.
- DeRisi, J.L., Iyer, V.R., and Brown, P.O. 1997. Exploring the metabolic and genetic control of gene expression on a genomic scale. *Science* 278: 680-686.
- Desprez, T., Amselem, J., Caboche, M., and Hofte, H. 1998. Differential gene expression in *Arabidopsis* monitored using cDNA arrays. *Plant J.* 14: 643-652.
- Diatchenko, L., Lau, Y.F.C., Campbell, A.P., Chenchik, A., Moqadam, F., Huang, B., Lukyanov, S., Lukyanov, K., Gurskaya, N., Sverdlov, E.D., and Siebert, P.D. 1996. Suppression subtractive hybridization: A method for generating differentially regulated or tissue-specific cDNA probes and libraries. *Proc. Natl. Acad. Sci. USA* 93: 6025-6030.
- Dudareva, N., D'Auria, J.C., Nam, K.H., Raguso, R.A., and Pichersky, E. 1998. Acetyl-CoA:benzylalcohol acetyltransferase - an enzyme involved in floral scent production in *Clarkia breweri*. *Plant J.* 14: 297-304.
- Duggan, D.J., Bittner, M., Chen, Y.D., Meltzer, P., and Trent, J.M. 1999. Expression profiling using cDNA microarrays. *Nature Genet.* 21 Suppl.: 10-14.
- Eisen, M.B., Spellman, P.T., Brown, P.O., and Botstein, D. 1998. Cluster analysis and display of genome-wide expression patterns. *Proc. Natl. Acad. Sci. USA* 96: 10943.
- Fodor, S.P.A., Read, J.L., Pirrung, M.C., Stryer, L., Lu, A.T., and Solas, D. 1991. Light-directed, spatially addressable parallel chemical synthesis. *Science* 251: 767-773.
- Fujii, T., Nagasawa, N., Iwamatsu, A., Bogaki, T., Tamai, Y., and Hamachi, M. 1994. Molecular cloning, sequence analysis, and expression of the yeast alcohol acetyltransferase gene. *Appl. Environ. Microbiol.* 60: 2786-2792.
- Fujii, T., Yoshimoto, H., Nagasawa, N., Bogaki, T., Tamai, Y., and Hamachi, M. 1996. Nucleotide sequences of alcohol acetyltransferase genes from lager brewing yeast, *Saccharomyces carlsbergensis*. *Yeast* 12: 593-598.
- Girke, T., Todd, J., Ruuska, S., White, J., Benning, C., and Ohlrogge, J. 2000. Microarray analysis of developing *Arabidopsis* seeds. *Plant Physiol.* 124: 1570-1581.

- Given, N.K., Venis, M.A., and Grierson, D. 1988. Hormonal regulation of ripening in the strawberry, a non-climacteric fruit. *Planta* 174: 402-406.
- Harada, M., Ueda, Y., and Iwata, T. 1985. Purification and some properties of alcohol acetyltransferase from banana fruit. *Plant Cell Physiol.* 26: 1067-1074.
- Harmer, S.L., Hogenesch, L.B., Straume, M., Chang, H.S., Han, B., Zhu, T., Wang, X., Kreps, J.A., and Kay, S.A. 2000. Orchestrated transcription of key pathways in *Arabidopsis* by the circadian clock. *Science* 290: 2110-2113.
- Harpster, M.H., Brummell, D.A., and Dunsmuir, P. 1998. Expression analysis of a ripening-specific, auxin-repressed endo-1, 4-beta-glucanase gene in strawberry. *Plant Physiol.* 118: 1307-1316.
- Hertzberg, M., Sievertzon, M., Aspeborg, H., Nilsson, P., Sandberg, G., and Lundeberg, J. 2001. cDNA microarray analysis of small plant tissue samples using a cDNA tag target amplification protocol. *Plant J.* 25: 585-591.
- Honkanen, E., and Hirvi, T. 1990. The flavour of berries. *Dev. Food Sci.* Amsterdam: Elsevier Scientific Publications.
- Jiang, M., Ryu, J., Kiraly, M., Duke, K., Reinke, V., and Kim, S.K. 2001. Genome-wide analysis of developmental and sex-regulated gene expression profiles in *Caenorhabditis elegans*. *Proc. Natl Acad. Sci. USA* 98: 218-223.
- Kane, M.D., Jatkoe, T.A., Stumpf, C.R., Lu, J., Thomas, J.D., and Madore, S.J. 2000. Assessment of the sensitivity and specificity of oligonucleotide (50mer) microarrays. *Nucl. Acids Res.* 28: 4552-4557.
- Lander, E.S. 1999. Array of hope. *Nature Genet.* 21 Suppl.: 3-4.
- Lashkari, D.A., DeRisi, J.L., McCusker, J.H., Namath, A.F., Gentile, C., Hwang, S.Y., Brown, P.O., and Davis, R.W. 1997. Yeast microarrays for genome wide parallel genetic and gene expression analysis. *Proc. Natl Acad. Sci. USA* 94: 13057-13062.
- Lipshutz, R.J., Fodor, S.P.A., Gingeras, T.R., and Lockhart, D.J. 1999. High density synthetic oligonucleotide arrays. *Nature Genet.* 21 Suppl.: 20-24.
- Lockhart, D.J., Dong, H., Byrne, M.C., Follettie, M.T., Gallo, M.V., Chee, M.S., Mittmann, M., Wang, C., Kobayashi, M., Horton, H., and Brown, E.L. 1996. Expression monitoring by hybridization to high-density oligonucleotide arrays. *Nature Biotech.* 14: 1675-1680.
- Maarse, H. 1991. Volatile compounds in foods and beverages. Marcel Dekker, Inc., New York.
- Maleck, K., Levine, A., Eulgem, T., Morgan, A., Schmid, J., Lawton, K.A., Dangel, J.L., and Dietrich, R.A. 2000. The transcriptome of *Arabidopsis thaliana* during systemic acquired resistance. *Nature Genet.* 26: 403-410.
- Manning, K. 1994. Changes in gene expression during strawberry fruit ripening and their regulation by auxin. *Planta* 194: 62-68.
- Manning, K. 1998. Isolation of a set of ripening-related genes from strawberry: their identification and possible relationship to fruit quality traits. *Planta* 205: 622-631.
- Mir, K.U., and Southern, E.M. 1999. Determining the influence of structure on hybridization using oligonucleotide arrays. *Nature Biotech.* 17: 788-792.
- Moyano, E., Portero-Robles, I., Medina-Escobar, N., Valpuesta, V., Munoz-Blanco, J., and Caballero J.L. 1998. A fruit-specific putative dihydroflavonol 4-reductase gene is differentially expressed in strawberry during the ripening process. *Plant Physiol.* 117: 711-716.
- Okamoto, T., Suzuki, T., and Yamamoto, N. 2000. Microarray fabrication with covalent attachment of DNA using Bubble Jet technology. *Nature Biotech.* 18: 438-441.

- Olias, J.M., Sanz, C., Rios, J.J., and Perez, A.G. 1995. Substrate specificity of alcohol acyltransferase from strawberry and banana fruits. ACS symposium series.
- Perez, A.G., Rios, J.J., Sanz, C., and Olias, J.M. 1992. Aroma components and free amino acids in strawberry variety Chandler during ripening. J. Agric. Food Chem. 40: 2232-2235.
- Perez, A.G., Sanz, C., and Olias, J.M. 1993. Partial purification and some properties of alcohol acyltransferase from strawberry fruits. J. Agric. Food Chem. 41: 1462-1466.
- Perez, A.G., Sanz, C., Olias, R., Rios, J.J. and Olias, J.M. 1996. Evolution of strawberry alcohol acyltransferase activity during fruit development and storage. J. Agric. Food Chem. 44: 3286-3290.
- Perez, A.G., Sanz, C., Olias, R., and Olias, J.M. 1999. Lipxygenase and hydroperoxide lyase activities in ripening strawberry fruits. J. Agric. Food Chem. 47: 249-253.
- Reymond, P., Weber, H., Damond, M., and Farmer E.E. 2000. Differential gene expression in response to mechanical wounding and insect feeding in *Arabidopsis*. Plant Cell 12: 707-719.
- Roscher, R., Schreier, P., and Schwab, W. 1997. Metabolism of 2,5-dimethyl-4-hydroxy-3(2H)-furanone in detached ripening strawberry fruits. J. Agric. Food Chem. 45: 3202-3205.
- Ruan, Y., Gilmore, J., and Conner, T. 1998. Towards Arabidopsis genome analysis: Monitoring expression profiles of 1400 genes using cDNA microarrays. Plant J. 15: 821-833.
- Schaffer, R., Landgraf, J., Accerbi, M., Simon, V., Larson, M., and Wisman, E. 2001. Microarray analysis of diurnal and circadian-regulated genes in *Arabidopsis*. Plant Cell 13: 113-123.
- Scharpf, L.G.Jr., and Chandan, R.C. 1989. Biotechnology and natural flavours for yogurt. National Yogurt Association, Virginia, USA.
- Schena, M., Shalon, D., Davis, R.W., and Brown, P.O. 1995. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. Science 270: 467-470.
- Schenk, P.M., Kazan, K., Wilson, I., Anderson, J.P., Richmond, T., Somerville, S.C., and Manners, J.M. 2000. Coordinated plant defense responses in *Arabidopsis* revealed by microarray analysis. Proc. Natl Acad. Sci. USA 97: 11655-11660.
- Schwab, W. 1998. Application of stable isotope ratio analysis explaining the bioformation of 2,5-dimethyl-4-hydroxy-3(2H)-furanone in plants by a biological Maillard reaction. J. Agric. Food Chem. 46: 2266-2269.
- Seki, M., Narusaka, M., Abe, H., Kasuga, M., Yamaguchi-Shinozaki, K., Carninci, P., Hayashizaki, Y., and Shinozaki, K. 2001. Monitoring the expression pattern of 1300 *Arabidopsis* genes under drought and cold stresses by using a full-length cDNA microarray. Plant Cell 13: 61-72.
- Seymour, G.B., Taylor, J.E., and Tucker, G.A. 1993. Biochemistry of fruit ripening. Chapman & Hall, London.
- Shimkets, R.A., Lowe, D.G., Tai, J.T., Sehl, P., Jin, H., Yang, R., Predki, P.F., Rothberg, B.E., Murtha, M.T., Roth, M.E., Shenoy, S.G., Windemuth, A., Simpson, J.W., Simons, J.F., Daley, M.P., Gold, S.A., McKenna, M.P., Hillan, K., Went, G.T., and Rothberg, J.M. 1999. Gene expression analysis by transcript profiling coupled to a gene database query. Nature Biotech. 17: 798-803.
- St-Pierre, B., Laflamme, P., Alarco, A.M., De-Luca, V. 1998. The terminal O-acetyltransferase involved in vindoline biosynthesis defines a new class of proteins responsible for coenzyme A-dependent acyl transfer. Plant J. 14: 703-713.

- Ueda, Y., Tsuda, A., Bai, J.H., Fujishita, N., and Chachin, K. 1992. Characteristic pattern of aroma ester formation from banana, melon, and strawberry with reference to the substrate specificity of ester synthetase and alcohol contents in pulp. *J. Jpn. Soc. Food Sci. Technol.*: 183-187.
- Velculescu, V.E., Zhang, L., Vogelstein, B., and Kinzler, K.W. 1995. Serial analysis of gene expression. *Science* 270: 484-487.
- Wang, C.L., Chin, C.K., Ho, C.T., Hwang, C.F., Polashock, J.J., and Martin, C.E. 1996. Changes of fatty acids and fatty acid-derived flavour compounds by expressing the yeast Delta-9 desaturase gene in tomato. *J. Agric. Food Chem.* 44: 3399-3402.
- Wang, R., Guegler, K., LaBrie Samuel, T., and Crawford N.M. 2000. Genomic analysis of a nutrient response in *Arabidopsis* reveals diverse expression patterns and novel metabolic and potential regulatory genes induced by nitrate. *Plant Cell* 12: 1491-1509.
- Welford, S.M., Gregg, J., Chen, E., Garrison, D., Sorensen, P.H., Denny, C.T., and Nelson, S.F. 1998. Detection of differentially expressed genes in primary tumor tissues using representation differences analysis coupled to microarray hybridization. *Nucl. Acids Res.* 26: 3059-3065.
- Wintoch, H., Krammer, G., and Schreier, P. 1991. Glycosidically bound aroma compounds from two strawberry fruit species, *Fragaria vesca* f. *semperflorens* and *Fragaria ananassa*, cultivar Korona. *Flavour Fragrance J.* 6: 209-216.
- Wisman, E., and Ohlrogge, J. 2000. *Arabidopsis* microarray service facilities. *Plant Physiol.* 124: 1468-1471.
- Woodward, J.R. 1972. Physical and chemical changes in developing strawberry fruits. *J. Sci. Food Agric.* 23: 465-473.
- Yamakawa, Y., Goto, S., and Yokotsuka, I. 1978. Fractionation and some properties of acetic-ester synthesizing enzyme from *Cladosporium cladosporioides* No.9. *Agric. Biol. Chem.* 42: 269-274.
- Yang, G.P., Ross, D.T., Kuang, W.W., Brown, P.O. and Weigel, R.J. 1999. Combining SSH and cDNA microarrays for rapid identification of differentially expressed genes. *Nucl. Acids Res.* 27: 1517-1523.
- Yoshioka, K., and Hashimoto, N. 1984. Ester formation by brewers' yeast during sugar fermentation. *Agric. Biol. Chem.* 45: 333-340.
- Zabetakis I and Holden MA (1997) Strawberry flavour: analysis and biosynthesis. *J. Sci. Food Agric.* 74: 431-434.
- Zhu, T. and Wang, X. 2000. Large-scale profiling of the *Arabidopsis* transcriptome. *Plant Physiol.* 124: 1472-1476.
- Zou, S., Meadows, S., Sharp, L., Jan, L.Y. and Jan, Y.N. 2000. Genome-wide study of aging and oxidative stress response in *Drosophila melanogaster*. *Proc. Natl Acad. Sci. USA* 97: 13726-13731.

Sandeep Kumar* and Matthias Fladung

*BFH Institute for Forest Genetics and Forest Tree Breeding, Sieker Land Str. 2, 22927 Grosshansdorf, Germany. *E-mail: kumar@holz.uni-hamburg.de*

1. INTRODUCTION

The production of transgenic plants is increasingly becoming an important component of agricultural biotechnology. For commercial success of plant transgenics, however, the questions need to be addressed which are related to the stable integration of transgenes and faithful transmission of introduced traits through successive generations in predictable manner. The transgene incorporated into the plant genome is integrated randomly and in unpredictable copy numbers, often in the form of repeats abolishing the expression of transgene (Stam et al. 1997, Kohli et al. 1999, Kooter et al. 1999, Kumar and Fladung 2001a) and sometimes may lead to excision of the transgene (Fladung 1999). The integration site also has a profound effect on expression of the transgene which is affected by inherent and extrinsic factors that may trigger methylation and reduce stability of the expression (Kumar and Fladung 2001a). Furthermore, insertion of the transgene in or near another gene may cause an undesired phenotype.

Targeting a single copy of a transgene into a pre-determined plant genomic location provides an efficient tool for securing long term stable expression. A gene targeting protocol may include: a) the ability to insert the transgene into regions of chromatin compatible with a desired developmental and tissue specific expression, and avoiding the knock-out of endogenous genes, b) the ability to control copy number, and/or c) subsequent removal of transgene repeats, selection marker genes, and other undesired DNA sequences. Targeted transgenes should therefore provide a more efficient and informative means of securing and comparing the expression of various transgenic sequences than is available with current transgenic procedures. Applications of gene targeting may also be extended to the inactivation of undesired traits, modification of existing genes, and elucidation of gene function through the analysis of gene malfunction. As the sequencing projects are progressing in plants, the creation of null mutants by targeting of specific sequences would be a direct way of relating sequences to the genes, and subsequently genes to phenotypes.

The process of transgene integration has important implications on gene targeting. Therefore, before describing gene targeting strategies in plants it is worthwhile to discuss the mechanisms of transgene integration. Since *Agrobacterium*-mediated T-DNA (transferred DNA) transfer is well established and characterised in model plants like *Arabidopsis* and tobacco, the process and mechanism of *Agrobacterium* T-DNA integration will be described briefly.

2. T-DNA INTEGRATION IN PLANTS

Agrobacterium tumefaciens is the most widely utilised vector for genetic transformation of plants. It contains a plasmid called Ti-(tumor inducing) plasmid (Figure 1) which encodes most of the major functions required for virulence (Bevan and Chilton 1982). *Agrobacterium* induces neoplastic growth in many plant species by transporting a single-stranded version of the T-DNA into the plant genome (Dumas et al. 2001). The T-DNA resides on the Ti-plasmid and is defined and delimited by 25-bp direct border repeats at its ends. While the wild-type T-DNA carries Ti genes, any DNA placed between the T-DNA borders will be transferred into the host cell and imported into its nucleus (Ballas and Citovsky, 1997).

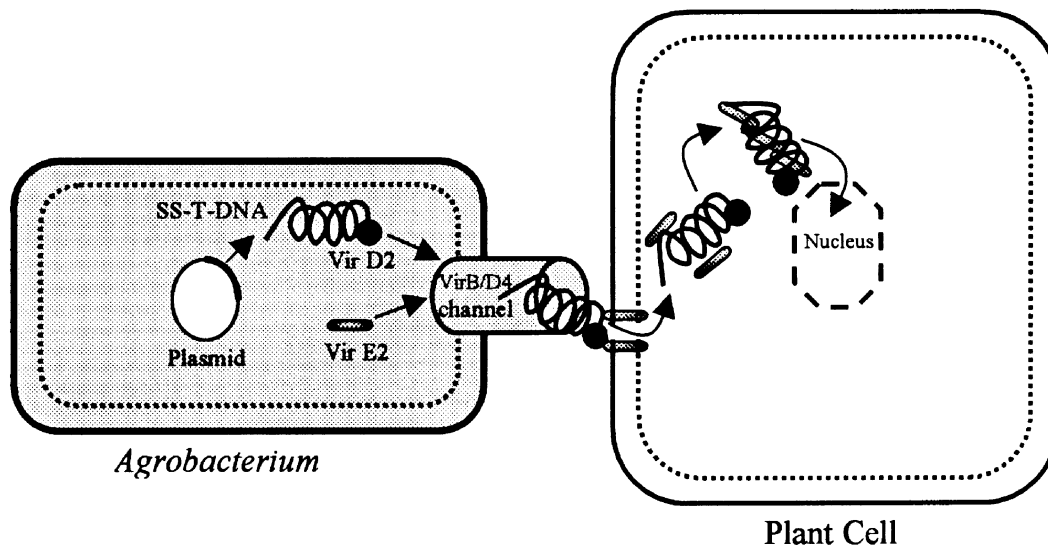


Figure1. Schematic diagram of the T-DNA transfer from Agrobacterium to the plant cell. With the induction of vir genes, ss-T-DNA is excised from the plasmid. The ss-T-DNA is processed with Vir proteins which bind the ss-T-DNA forming a T-DNA complex. The T-DNA complex is transferred from the bacterial to the plant cell via Vir proteins-made membrane channels. In the plant cell the ss-T-DNA is protected by VirE2 from nucleolytic attacks and VirD2 pilots the T-DNA complex through the nuclear pore (Dumas et al. 2001).

2.1. T-DNA TRANSFER FROM BACTERIA TO THE PLANT CELL

In the presence of specific small phenolic molecules released by wounded plants, *Agrobacterium* initiates gene expression from the Ti-plasmid virulence (Vir) region, which produces most of the trans-acting factors for

T-DNA transfer. Products of *vir* genes act at the T-DNA borders to generate a linear single stranded copy of the T-DNA (Zupan et al. 1996). VirD2, with the help of VirD1, recognises and cleaves within two 25-bp border sequences. On cleavage, VirD2 binds covalently to the 5' end of the single-stranded T-DNA (ss-T-DNA), forming the T-DNA strand (Figure 1).

The VirE2 protein binds to ss-T-DNA and protects the T-DNA complex from nucleolytic attack. VirD2 and VirE2, together with ss-T-DNA are the only known essential components that are exported from the bacterial cell. VirE2 forms a membrane channel that transfers T-DNA complex through the plant plasma membrane, and once in plant cytoplasm VirD2 pilots the T-DNA complex through the nuclear pore (Dumas et al. 2001). Finally, the T-DNA integrates into the plant genome with the help of plant enzymes (Gelvin 2000).

2.2. CHARACTERISATION OF TRANSGENE INSERTION LOCUS

The foreign genes may be integrated into the host genome by two different ways: homologous recombination, in which a region of sequence identity between the partners is needed; and illegitimate recombination, requiring no sequence-specific homology. The controlled manipulations of genomes by homologous recombination is desirable, because the recombination products can be predicted in advance (Puchta, 1998). In order to get insight into the mechanism of transgene integration in plants transgene insertions sites in plant chromosomes have been characterised (Figure 2a). Nucleotide sequence analysis of the target sites was carried out in different *Nicotiana*, *Arabidopsis*, and *Populus* transgenic plants lines before and after the integration (Gheysen et al. 1991, Matsumoto et al. 1990, Mayerhofer et al. 1991, Fladung 1999, Kumar and Fladung 2000a).

The nucleotide sequence comparison of wild type and T-DNA tagged genomic loci showed that junction formation with plant DNA may lead to short deletions at both ends of T-DNA (Figure 2b) as well as unpredictable deletions of host genomic sequences. In general two classes of T-DNA/plant junctions were observed. First, terminal sequences of T-DNA ends were found to show similarities with short segments of target plant DNA near the breakpoints resulting into precise target replacements. Second, imprecise T-DNA/plant junctions were observed and no sequence similarities was found between plant DNA and one of the T-DNA ends. Such junctions were accompanied with intervening filler DNA which in most cases resembled to the T-DNA or plant DNA near break points (Mayerhofer et al. 1991). The position of the base-paired segments between T-DNA and target DNAs seems to define the position of nicks in the target DNA (Koncz et al. 1994). Based on these studies in plants and in the light of results obtained in our lab on transgenic aspen-*Populus* (Kumar and Fladung 2000a) a simple generalised model was suggested for

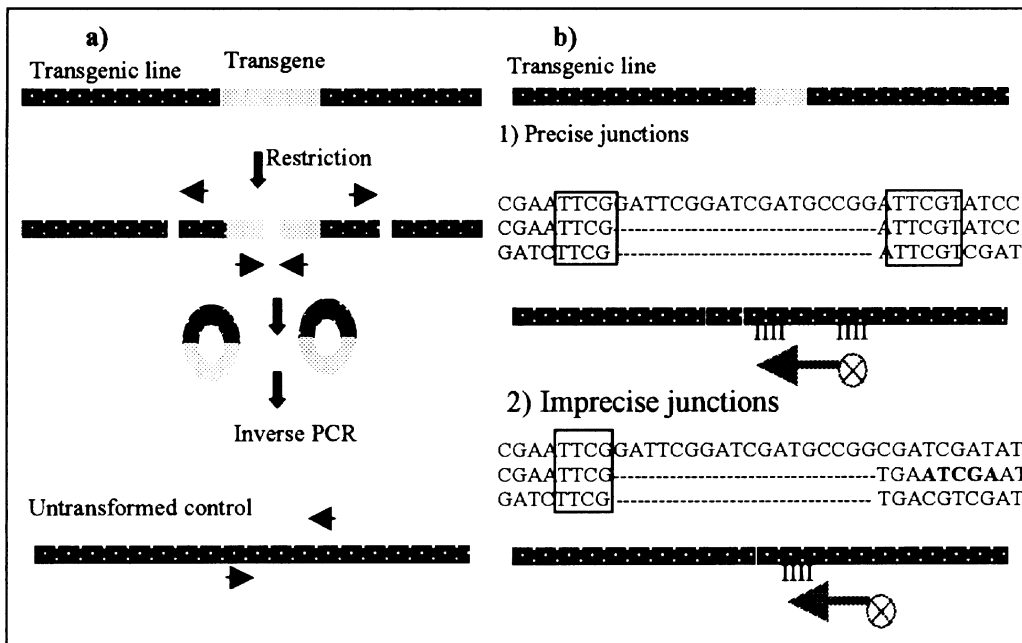


Figure 2. Mapping of the transgene integration site and T-DNA integration model. **a)** Genomic DNA from the transgenic line is digested with a restriction enzyme having single restriction site in the transgene. The restriction is followed by self-ligation resulting into two circular fragments containing part of the transgene and the genomic regions flanking left and right borders of the transgene. These circular fragments could be amplified with PCR using inverse primers (shown with filled arrows) followed by sequencing to determine the flanking genomic regions. Primers could be designed from the flanking genomic regions to amplify the transgene integration site from non-transformed control. **b)** A hypothetical comparison of the transgene integration site before (upper sequence) and after (middle sequence) T-DNA integration, lower sequence represents the transgene. Two classes of integration sites are generally observed: 1), precise junctions showing small sequence identity (see boxes) at junctions, and 2), imprecise junctions which do not show sequence identity at least for one junction which is comprised of short filler sequence (bold letters). A T-DNA integration model has been suggested (Kumar and Fladung 2000a), according to which 3'-end of the transgene (arrowhead) search for sequence similarity (shown as IIII) in the plant DNA and VirD2-linked-5'-end (circled cross) anneals transiently to the complimentary chromosomal strand. When 5'-end is also able to find sequence similarity, precise junctions are formed. In the event of no sequence similarity between the 5'-end of invading T-DNA strand and plant DNA, imprecise junctions are formed because of blunt ligation of the 5'-end to the plant DNA resulting into filler formation.

transgene integration in plants (Figure 2b). However, keeping in view the complex nature of transgene integration phenomenon in plants, the model may not explain all the transgene integration events taking place in plants.

2.3. MECHANISM OF T-DNA INTEGRATION

In majority of simple T-DNA insertions without any repeats microhomologies between T-DNA and the target DNA associated with short deletions in T-DNA have been found only at 3'-end of the invading T-DNA. The 5'-end of the integrated T-DNA is usually conserved and

contains three nucleotides from the 25-bp T-DNA border repeat (Koncz et al. 1994). Similar observations were made while analysing complex integration loci in transgenic aspen-*Populus* containing T-DNA repeats (Kumar and Fladung 2000a). The 5'-end of the T-DNA is tightly associated with the VirD2 protein which pilots the T-DNA complex through the nuclear pore.

Therefore, it is suggested that during the T-DNA integration process the 3'-end of the invading T-DNA strand finds the short sequence similarity in the plant DNA resulting into endonucleolytic digestion of the displaced plant DNA. The overhanging 3'-end of the T-DNA beyond microhomology is digested exonucleolytically or by an endonuclease cleaving (Tinland and Hohn 1995) explaining the T-DNA deletions at 3'-end. The 5'-end linked to VirD2 scans for the sequence similarity in the complementary target strand. In the event of short-similarity-based integration of both T-DNA ends precise T-DNA/plant junctions will be formed without any intervening DNA or fillers. However, it is conceivable that the 5'-end of the T-DNA fails to find short sequence similarity in the complementary target strand. Under these circumstances the nucleotide attached to the VirD2 may anneal to the plant strand. Since the 5'-end is ligated without any sequence similarity to the complementary plant strand, unrelated free 3'-ends may be used as primers for limited copying to fill the gap. These 3'-ends may be available from the nicked T-DNA strand or from the host DNA, thus explaining the resemblance between the fillers and the sequences near to the break points in T-DNA strand or the host integration sites (Kumar and Fladung 2000a).

These observations on T-DNA integration in plants suggest that the mechanism by which T-DNA integrates into plant chromosome is non-homologous or illegitimate recombination. It means that foreign DNA integrates randomly into the plant genome, regardless of whether an homologous region is included in the targeting sequence. This is in contrary to the transgene integration in prokaryotes and lower eukaryotes where homologous recombination is the predominant way of recombination and homologous recombination-based gene targeting is a routine. In yeast, foreign DNA harbouring sequences with identity to genomic sequences almost exclusively integrate into the corresponding genomic position. Transformation with sequences that carry no homology to the yeast genome is rarely successful (Puchta 1998).

3. HOMOLOGOUS RECOMBINATION-BASED GENE TARGETING

Homologous recombination involves the exchange of covalent linkages between DNA molecules in the regions of highly similar or identical sequence. An important aspect of homologous recombination is its possible application in manipulating the plant genome. It can be used to target a

transgene to any chosen locus in the genome. The establishment of site-directed gene-targeting techniques in the mouse embryonic stem cells has opened a new chapter in our knowledge of the mammalian biology. Such a gene-targeting system is highly desirable in crop plants and some attempts have been made in tobacco and *Arabidopsis* using either direct gene transfer techniques or *Agrobacterium*-mediated transformation. Selectable markers were used as model substrate in most of these experiments. The gene targeting rates obtained were as low as one targeted event in 10^4 - 10^6 transformation events, suggesting the strong prevalence of illegitimate recombination over homologous recombination (Puchta and Hohn 1996).

Studies in prokaryotes and lower eukaryotes have provided much insight into the nature of this process, the recombination intermediates, the genes, and the proteins involved (Shalev et al. 1999). Homologous recombination has been studied in plants and models for meiotic and somatic recombination have been suggested (Puchta and Hohn 1996). Our improved understanding of homologous recombination has so far not resulted in any significant increase in targeted integration events in plants (Puchta 1998). Gene targeting of artificially created genomic targets has been attempted in plants through direct delivery of DNA (Paszkowski et al. 1988, Hafter et al. 1992), or by *Agrobacterium* infection (Lee et al. 1990, Offringa et al. 1990). Although, some success in gene targeting has been reported in *Arabidopsis* via *Agrobacterium* infiltration *in planta* (Miao and Lam 1995, Kempin et al. 1997), number of targeting events was unfortunately so small and variable that it is not possible to obtain a statistically valid conclusion. Despite of the extended length of homology in the transferred DNA (Miao and Lam 1995, Risseuw et al. 1997, Thykjaer et al. 1997) the predominant mechanism of transgene integration in higher plants still remains illegitimate recombination which involves very small homology between recombining DNA strands.

Some alternative approaches have been suggested and applied to obtain gene targeting in plants (Kumar and Fladung 2001b). First, the inability to perform homologous recombination in plants may be complemented by engineering the genome of higher plants by homologous recombination-related genes from other organisms (Puchta and Hohn 1996). The second approach may be the activation of the target locus and/or donor DNA by inducing double strand break (DSB) via site-specific endonucleases combined with site-specific recombination system. The third possibility is to determine the suitable plant tissues and target the stage when they are active for homologous recombination and gene targeting.

4. TRANSFER OF HOMOLOGOUS RECOMBINATION-RELATED GENE FROM OTHER SPECIES

As homologous recombination occurs efficiently in many lower eukaryotes such as budding yeast and bacteria, gene targeting has proved to be a very powerful tool in these organisms. The process of homologous recombination requires search for homology, recognition of sequence similarity, and strand exchange between two DNA molecules. In *Escherichia coli* these different steps together with the proteins required are: initiation of homologous recombination by a DNA double-strand break and/or single-strand DNA formation by the RecBCD complex; exchange of DNA strands, including homologous recognition and strand displacement, done by RecA-like proteins; heteroduplex extension, performed by RuvA, RuvB or RecG to yield a recombination intermediate, named the Holliday junction; and resolution of the heteroduplex Holliday junction by the endonuclease RuvC (Shalev et al. 1999). The RecA protein plays a central role in the recombination pathway of *E. coli*. A number of proteins with similarity to RecA have been found in higher eukaryotic cells including plants.

The distinguishing features of most of these proteins is that they do not require ATP to stimulate strand exchange. Rather than a search for homology and promotion of strand invasion, recombination in plant involves exposure of single-stranded DNA by exonuclease and subsequent reformation of double-stranded DNA from complementary single-stranded DNA. Hence the mechanism is inherently different from that of RecA-mediated recombination. Therefore, several RecA-like proteins seem to cooperate in plant system and other class of strand exchange proteins might be needed at the same time (Reiss et al. 1996).

A possible alternative is to express RecA, the key protein of the *E. coli* recombination pathway in plants to induce homologous recombination. By using this approach (Reiss et al. 1996), it was shown in *recA* transgenic tobacco plants that the prokaryotic nucleus-targeted RecA protein by itself stimulated genomic homologous recombination in the plant cells. The observations suggested that the single protein was sufficient to perform the search for homology and strand exchange in a higher eukaryotic cell. Similarly, transgenic tobacco plants expressing RuvC were associated with a 12-fold increase in somatic crossover between genomic sequences, with an 11-fold increase in intrachromosomal recombination, and with a 56-fold increase in extrachromosomal recombination between plasmids cotransformed into leaves via particle bombardment (Shalev et al. 1999). The mechanism by which these proteins stimulates homologous recombination in plants is not clear, but the expression of these proteins can lead to the engineering of recombinogenic plants. Such plants, in combination with the development of improved gene targeting vectors, may enable high frequencies of exogenous DNA integration into

chromosomal target via homologous recombination thus facilitating gene targeting in plants.

5. INDUCING DOUBLE STRANDED BREAKS (DSBs) VIA SITE-SPECIFIC ENDONUCLEASES

Genomic double strand breaks (DSBs) are the key intermediates in the recombination reactions of living organisms and play an important role in homologous recombination in eukaryotes. Yeast mitochondrial endonuclease *I-Sce* I (Perrin et al. 1993), which has an 18-bp recognition site has been used to induce genomic DSBs and their repair by homologous recombination in mouse cells (Rouet et al. 1994, Chouluka et al. 1995). In plants, factors that induce unspecific DSBs in DNA such as X-rays (Lebel et al. 1993) or methyl methanesulfonate (Puchta et al. 1995) were shown to enhance homologous recombination. It has also been shown that *I-Sce* I-induced site-specific genomic DSBs in plant cells were repaired by homologous recombination by exogenously supplied DNA (Figure 3). Induction of the DSB led to drastically enhanced frequencies of

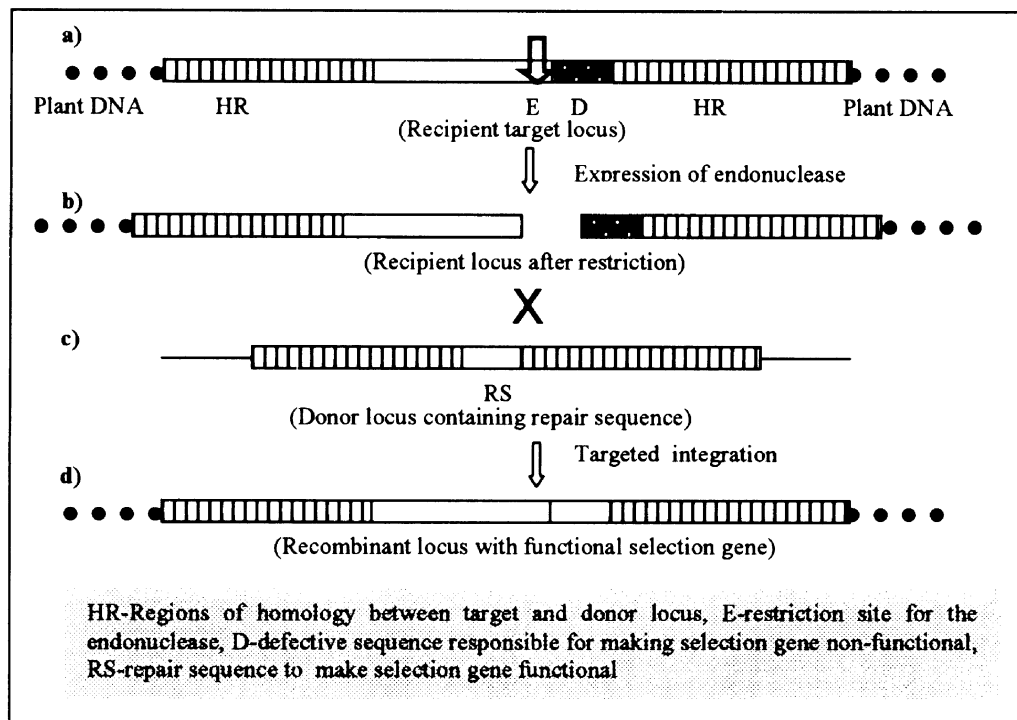


Figure 3. Gene targeting based on site-specific endonuclease. **a)** Recipient target locus in the plant genome containing the identical sequence (HR) to donor sequence, endonuclease restriction site (E), defect (D) in the selection gene making it non-functional. **b)** Break in the recipient locus after expression of the endonuclease. Donor locus containing sequence identical (shaded boxes) to the target recipient locus and a repair sequence (RS). **d)** Targeted integration will lead to replacement of defective sequence (D) by repair sequence (RS) making selection gene functional, and the targeted events could be screened on the selection medium (Puchta et al. 1993).

homologous integration into specific locus of the plant chromosome (Puchta et al. 1996). The induction of a DSB might in general be one of the rate-limiting steps in gene targeting in plants. Thus a technique based on the induction of restriction-mediated DSBs might be established for gene targeting in plants.

Another important aspect determining the overall efficiency of gene targeting may be the recombinogenic status of the donor DNA. In organisms in which gene targeting has been achieved, donor DNA molecules with cut or broken ends have proven to be recombinogenic. Puchta et al. (1993) have demonstrated that the expression of the *I-Sce I* gene led to specific *in vivo* DSBs in DNA molecules transfected into plant protoplasts. As a consequence extrachromosomal homologous recombination was induced. To repair a DSB by homologous recombination, a donor sequence is required which is homologous to the region to be repaired. Because the chromatids are identical to each other they might be the preferred substrate for the DSB repair over extrachromosomal donor DNA.

Therefore, in future a gene targeting strategy should be designed that produces an extrachromosomal recombinogenic donor molecule which is able to compete with sister chromatids to repair the DSBs in order to increase the number of the targeted events (Kumar and Fladung 2001b). One factor may be the extent of the homology or non-homology between the donor and the target. Another approach to obtain homology-directed changes in a target locus may be to use broken-ended extrachromosomal DNA molecules. Recently such a method for gene targeting has been designed in *Drosophila* that uses broken-ended extrachromosomal DNA molecules to produce homology-directed changes in a target locus (Rong and Golic 2000).

6. TARGETING DURING ACTIVE HOMOLOGOUS RECOMBINATION STAGES

Gene targeting and homologous recombination remains inefficient in plants although it has been used for two decades in bacteria and yeast, and has now successfully been applied in mammals and *Physcomitrella* (Reski 1998). Keeping this in mind it will be worthwhile to compare crop plants to the organisms known for the homologous recombination. Contrary to the seed plants *Physcomitrella* remains in haploid phase for the most of its life cycle. The other important factor for successful gene targeting in *Physcomitrella* may be the tissue used for transformation, which was not only haploid but also gametophytic (Reski 1998). It has, therefore, been suggested to obtain gene targeting using the gametophytic tissues of the seed plants. This can be tested as regeneration protocols for such tissues are available for many plant species. The successful knocking out of *AGL5* MADS-box gene in *Arabidopsis* (Kempin et al. 1997) has been later

speculated as a result of transformation of meristematic or meiotic tissues (Puchta 1998).

Similarly in mammals, most of the gene targeting success has been obtained from the mouse embryonic stem cells. In many animal systems it has been shown that genomes are packaged into the compartments that are either transcriptionally competent euchromatin or the repressive, transcriptionally silent heterochromatin. This compartmentalisation accompanied with DNA methylation arises progressively during early embryonic development (Roberston and Wolffe 2000). Interestingly, mouse embryonic stem cells known for homologous recombination can progress throughout the cell cycle and divide without detectable DNA methylation (reviewed in Roberston and Wolffe 2000) indicating that the early embryonic chromatin structure is maintained in these cells without any notable changes. It has, therefore, been speculated that embryonic tissues in their early development may be a better stage for targeting in plants compared to the leaf tissues which have undergone several somatic divisions and differentiation (Kumar and Fladung 2001b). Analysis of *mim* (hypersensitive to methylmethane sulfonate, irradiation, and mitomycin C) mutants from *Arabidopsis* indicates a major influence of chromatin accessibility on the *in vivo* recombination frequency of chromosomal DNA (Mengiste *et al.* 1999, Hanin *et al.* 2000). Indirect evidence that chromatin structure plays a role in homologous recombination has come from the fungus *Ascobolus immersus*, in which it was shown that methylation of a known meiotic-recombination hotspot reduced the frequency of crossing-over with this region several-hundred-fold (Maloisel and Rossignol 1998). If this holds true then manipulation of the chromatin structure may become a critical factor in future gene targeting experiments. Nevertheless, further experimentation is required to confirm such a notion and it needs some more time when we are able to manipulate the chromatin structure to make it accessible for recombination. Until then gene targeting may be attempted during the plant stages that are more active in homologous recombination to increase the frequency of gene targeting in plants.

7. SITE-SPECIFIC RECOMBINATION SYSTEM

Even though homologous recombination efficiency is increased and/or DSBs are created artificially, a donor sequence is required to repair a DSB by homologous recombination which should be recombinogenic and homologous to the region to be repaired. The chromatids which are identical to each other might be the preferred substrate for the DSB repair over extrachromosomal donor DNA. The only way to make donor DNA more recombinogenic is to look for the *in vivo* excision and recombination of the extrachromosomal target DNA that is pre-integrated randomly in the plant genome.

A site-specific recombination system may be used for *in vivo* excision of the donor DNA sequence. The two components of the recombination system are the site specific recombinase and its recognition site. The Cre/*lox* (recombinase/recognition site) system of bacteriophage P1, the FLP/*FRT* system of the yeast 2 μ m plasmid, the modified Gin/*gix* system of bacteriophage Mu, and the R/*RS* system of the yeast plasmid pSR1 are such recombination systems which are shown to supported by the plant cells (Odell and Russell 1994; Lloyd and Davis 1994; Lyznik et al., 1993, 1995). The site-specific protein recognises a defined sequence, the recognition site that occurs in the boundaries of a segment of DNA. The two recognition sites in the direct orientation lead to the excision of the DNA sequence between them, whereas the recognition sites in inverted orientation result in an inversion of the intervening sequence (figure 4). The site-specific recombination system, therefore may be an important tool to obtain *in vivo* excision of the donor DNA in a gene targeting strategy.

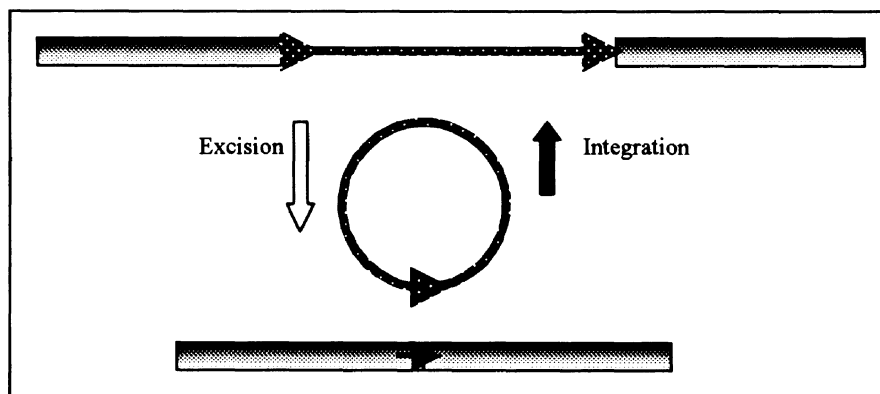


Figure 4. A site-specific recombination system. Excision of DNA segment between directly oriented recognition sites (filled arrowheads). The reverse reaction will cause integration. The lightly filled bar represents genomic DNA.

8. AN INTEGRATED GENE TARGETING STRATEGY

Different gene targeting methods described above have been tested in crop plants without much success. Although it has been shown that these methods are responsible for increased homologous recombination they may not be so efficient when extrachromosomal sequences are used for gene targeting. The process of homologous recombination or gene targeting in plants seems far more complex than our current level of the knowledge. One such example may be the recently published report on gene targeting (Reiss et al. 2000) in which homologous recombination-related transgene was used with and without artificially created DSBs. An artificially created locus was targeted in tobacco by overexpressing the recombination protein nuclear-targeted RecA (nt-RecA), relying on *Agrobacterium* transformation.

The gene targeting was assayed with two different artificial target loci including one containing site for the restriction enzyme *I-Sce* I. The number of homologous events observed were not significantly increased by the presence of nt-RecA even if recombination was induced by DSB. However, DSB induction revealed a major effect of nt-RecA on the distribution of recombination product classes and a significant increase in the ratio of true gene targeting events to ectopic events was observed. The authors suggest that stimulation of homologous recombination by RecA is limited to processes that transiently produce nascent single strands, as happens during DNA synthesis. Since the single stranded T-DNA is likely to be masked by the single-stranded-DNA-binding protein VirE2 from *Agrobacterium tumefaciens*, the authors presumed that RecA was inactive in this situation. These findings indicate that RecA cannot substantially improve gene targeting frequencies in plants when using the *Agrobacterium*-based transformation system.

Hence concerted efforts are required in future gene targeting experiments that combines expression of recombinase and endonuclease transgenes to the factors that induce *in vivo* nascent recombinogenic donor DNA sequence (Figure 5). Such a gene targeting strategy has already been suggested for plants (Kumar and Fladung 2001b) and need to be tested in model plants. This technique uses the endogenous DNA repair system of

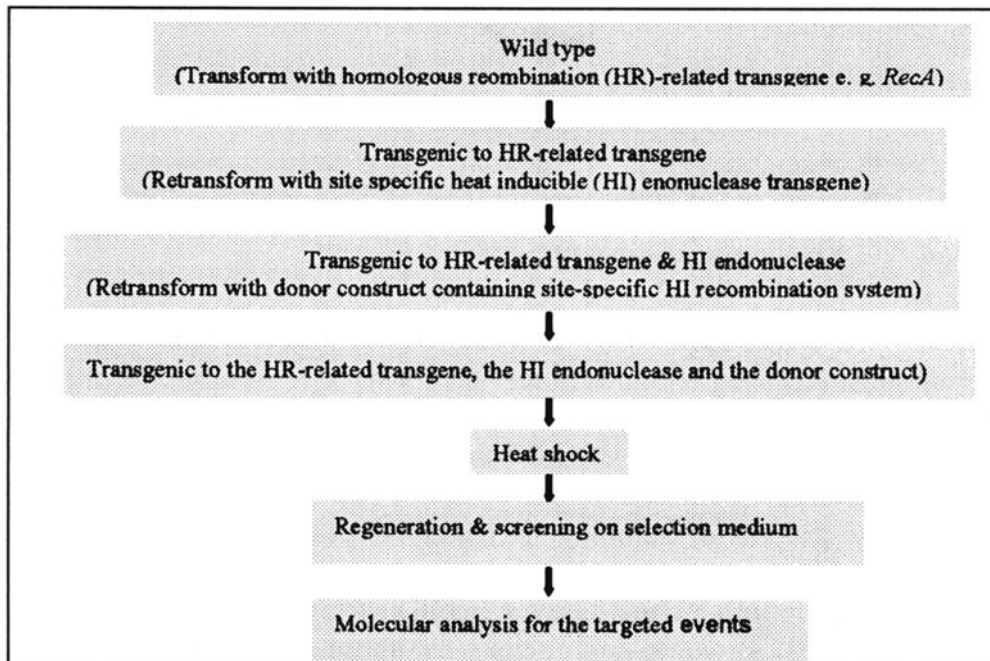
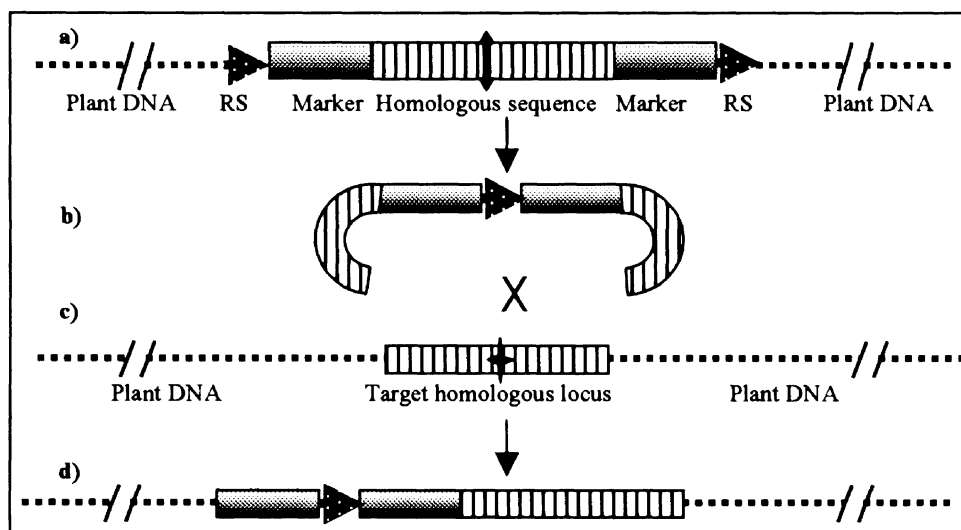


Figure 5. An integrated gene targeting strategy for plants. The scheme is based on series of transformations. However, it should be possible to obtain lines transgenic to the three required transgenes via co-transformation. The mechanisms of excision and recombination events happening after heat shock treatment are shown in Figure 6.

the host plant to integrate donor DNA at a target locus and combines almost all the factors that increase gene targeting or homologous recombination in plants.

The method comprises of three to four round of transformations to obtain plants that are transgenic to: the RecA recombinase gene, a heat inducible endonuclease gene, a heat inducible site-specific recombination system, and a donor DNA sequence that contains recognition sites for the site-specific recombinase, the endonuclease, and DNA sequence to be targeted. Heat shock treatment in the early stages of the development will result in a *in vivo* nascent recombinogenic donor sequence with broken ends and a functional recombined marker gene (Figure 6).

This gene targeting technique mainly provides a way to mutate genes to analyse their functions. The method needs to be improved to insert a transgene of interest into desired genomic position to obtain precisely tailored transgenic plants. For that purpose, it is important that a minimum



*Figure 6. Mechanisms of in vivo excision and targeted integration of extrachromosomal homologous donor DNA sequence. a) A donor sequence integrated randomly in the plant genome which is transgenic to all three transgenes required in the integrated gene targeting strategy (Figure 5). The donor sequence contains: directly oriented recognition sites (RS) of a site-specific recombination system (Figure 4), a non-functional selection marker sequence split into two parts (filled bars) as an excision marker, a homologous sequence (widely shaded bar) which is identical to the target locus but without defect, and a restriction site (double head filled arrow). b) The simultaneous induction of site-specific recombinase and endonuclease will result in *in vivo* excision of the donor sequence between two directly oriented recombination sites (RS) and restriction of the sequence homologous to the target (widely shaded bar), respectively. This will lead to the integration of the excision marker gene into a functional unit and splitting of the homologous sequence in two pieces with broken ends making it recombinogenic. c) A target locus containing a defective (filled star) non-functional sequence which is similar to the donor sequence. The targeted integration will lead to repair of the defective sequence which can be screened on the selection medium.*

number of transgenes are used to obtain targeted gene transfer events and all the foreign sequences other than the targeted sequence are subsequently removed from the host genome.

9. REMOVAL OF UNDESIRE DNA SEQUENCES

For the commercial success and for the public acceptance of transgenic plants, it is becoming increasingly important to remove the marker genes which are usually transferred with the gene of interest in order to select the putative transgenic lines. These marker genes become useless after the selection and require their removal to overcome the various public concerns. Further, the gene targeting strategy described here also uses a number of transgenes which need to be removed once the targeted integration is obtained. The undesirable DNA sequences can be eliminated by placing it between two directly orientated recognition sites of a site-specific recombinase followed by the expression of recombinase resulting in recombination between flanking recognition sites deleting the marker gene (Sugita et al. 2000, Puchta 2000). Since the gene targeting strategy described here uses such a recombination system, the marker genes and other sequences can subsequently be removed by the same system.

A similar but shortcut method based on intrachromosomal recombination between bacteriophage λ attachment (attP) regions has recently been developed to remove selectable marker genes from tobacco transgenes (Zubko et al 2000). This is a significant improvement over previous recombinase based methods since it does not require transfer of additional recombinase gene avoiding further work for the removal of recombinase gene. Such a tool may also be useful for the deletion of recombinase and endonuclease genes used in the suggested gene targeting strategy.

10. OBTAINING SINGLE COPY TRANSGENIC PLANTS

Single copy transgene integration is one of the important means to ensure the stable expression of the transgene. Although single copy transgenes might also be silenced (Elmayan and Vaucheret 1996, Meyer et al. 1992, Kumar and Fladung 2001a), a higher incidence of transgene instability is correlated with high transgene copy number (Assad et al. 1993, Atkinson et al. 1998, Meyer and Saedler 1996, Kumar and Fladung 2001a). Therefore, for long-term stable expression of a transgene, it would be critical to start with a single intact copy of the transgene. However, the occurrence of single copy integration is rare event when biolistic transformation method is used. Even with *Agrobacterium*-mediated transformation integration of more than one copy is a common phenomenon.

The conventional method of obtaining single copy transgenic plants is to screen among the pool of transformants using molecular methods. However, genetic transformation has become an important tool

not only for defining gene function but also for testing the commercial utility of new sequences. As the sequence information is increasing rapidly plant transgenics will also expand to reveal the functions of the unknown sequences. Therefore conventional screening methods need to be improved to handle growing number of transgenic plants. PCR-based quick methods to eliminate transgenic plants containing multiple copies of the transgene may be viable alternative for the initial screening of large number of transgenic plants (Kumar and Fladung 2000b, Spertini *et al.* 1999).

In addition to screening methods some efforts have been made to improve transformation procedures for obtaining increased number of single copy integration events. One such approach is the 'Agroloistic' transformation which combines *Agrobacterium* and particle bombardment-mediated technologies (Hansen and Chilton 1996, Hansen *et al.* 1997). This method is based on cobombarding *virD* genes along with T-DNA borders flanking the introduced transgenes. Srivastava *et al.* (1999) have used site-specific recombination-based strategy to obtain single-copy transgenic wheat through the resolution of complex integration patterns. The transformation vector contained a transgene flanked by recombination sites in an inverted orientation. Regardless of the number of copies integrated between the outermost transgenes, recombination between outermost sites resolved the integrated molecules into a single copy. It was shown that four out of four multiple-copy loci were resolved into single copy transgenes. Therefore, site-specific recombination system seems to be an indispensable tool for the second generation plant transgenics.

11. CHIMERIC RNA/DNA OLIGONUCLEOTIDES FOR TARGETED *IN VITRO* MANIPULATIONS

A technology currently being explored in prokaryotic and eukaryotic systems uses self-complementary chimeric oligonucleotides (COs) comprised of DNA and 2'-*O*-methyl RNA to target and mutate genes *in vivo*. The strategy exploits the natural recombinogenicity of RNA:DNA hybrids and feature double-hairpin capped ends avoiding destabilisation or destruction by cellular helicases or exonucleases (Yoon *et al.* 1996). These COs are designed to have one or more bases that do not pair with the endogenous gene sequences (Beetham *et al.* 1999). The molecule is a double stranded chimera consisting of DNA and RNA residues, capped at both ends by sequences which fold into a hairpin (see Cole-Strauss *et al.* 1999). The loops of four 'T' nucleotides are connected to a DNA-based sequence that is complementary to the target locus. A strand break in the lower strand allows topological interwinding of chimera into the target DNA. The O-methylation protects the RNA from degradation within cells (Hohn and Puchta 1999). The half-life of such chimeric structures is substantially longer than similar duplex molecules (Yoon *et al.* 1996). The chimera contains nucleotide(s) that differs from the target sequence, upon

binding, forms mismatched base(s) with the specific targeted nucleotide(s). The resulting helical distortion is recognised by the cell's DNA repair machinery and the base pair(s) corrected using the DNA sequence of the chimera as a template (Cole-Strauss et al. 1999).

This gene targeting method has been applied in tobacco (Beetham et al. 1999) and maize (Zhu et al. 1999). The results demonstrate that plant genes can be modified specifically and efficiently by COs and suggest that reverse genetics and engineering of endogenous genes in commercially important crops will be feasible by using this approach. However, the frequencies of site-specific targeting observed in plants are up to 3 orders of magnitude lower than the frequencies reported for COs-mediated nucleotide conversion in mammalian cells (Zhu et al. 1999). Delivery of COs is considered more difficult in plant cells than in animal cells because of the rigid plant cell wall (Hohn and Puchta 1999). This method involves alteration of 1-2 bp in the target site, which is useful for site-specific mutagenesis, gene knockouts, and allelic replacements. However, it may not compete with gene targeting based on homologous recombination for the wider application especially when larger changes are in question (Kumar and Fladung 2001b).

12. CONCLUDING REMARKS

The results in the recent literature indicate that some progress has been made in gene targeting in the plants and some preliminary success is in hands. It seems likely that some gene targeting methods will be available in the near future, at least for the reverse genetics point of view, to elucidate the gene function by mutating the endogenous genes. However, for plant transgenics prospective, a more comprehensive and integrated approach needs to be followed which includes manipulations of the host system and the donor sequence. The strategy suggested in this chapter may be a small step in this direction.

13. REFERENCES

- Assad F.F., Tucker K.L., Signer E.R. (1993) Epigenetic repeat-induced gene silencing (RIGS) in *Arabidopsis*. *Plant Mol. Bio.* 22, 1067-1085
- Atkinson R.G., Bielecki L.R.F., Gleave A.P., Jannsen B.-J., Morris B.A.M. (1998) Post-transcriptional silencing of chalcone synthase in petunia using a geminivirus-based episomal vector. *Plant J.* 15, 593-604
- Allen G.C., Hall G.E.J., Childs L.C., Weissinger A.K., Spiker S., Thompson W.F. (1993) Scaffold attachment regions increase reporter gene expression in stably transformed plant cells. *Plant Cell* 5, 603-613
- Ballas N., Citovsky V. (1997) Nuclear localization signal binding protein from *Arabidopsis* mediates nuclear import of *Agrobacterium* VirD2 protein. *Proc. Natl. Acad. Sci. USA* 94, 10723-10728

- Beetham, P. R., Kipp, P. B., Sawycky, X. L., Arntzen, C. J., May, G. D. (1999) A tool for functional plant genomics: Chimeric RNA/DNA oligonucleotides cause *in vivo* gene-specific mutations. *Proc. Natl. Acad. Sci. USA* 96, 8774-8778
- Bevan M.W., Chilton M.D. (1982) T-DNA of the *Agrobacterium* Ti and Ri plasmids. *Ann. Rev. Genet.* 16, 357-384
- Choulika A., Perrin A., Dujon B. and Nicolas J. F. (1995) Induction of homologous recombination in mammalian chromosomes by using the *I-SceI* system of *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* 15, 1968-1973
- Cole-Strauss, A., Gamper, H., Holloman, W. K., Munoz, M., Cheng, N., Kmiec, E. B. (1999) Targeted gene repair directed by the chimeric RNA/DNA oligonucleotide in a mammalian cell-free extract. *Nucleic Acids Res.* 27, 1323-1330
- Dumas F., Duckely M., Pelczar, P., Van Gelder P. and Hohn B. (2001) An *Agrobacterium* VirE2 channel for transferred-DNA transport into plant cells. *Proc. Natl. Acad. Sci. USA* 98, 485-490
- Elmayan T., Vaucheret H. (1996) Expression of single copies of a strongly expressed 35S transgene can be silenced post-transcriptionally *Plant J.* 9, 787-797
- Fladung, M. (1999) Gene stability in transgenic aspen (*Populus*). I. Flanking DNA sequences and T-DNA structure. *Mol. Gen. Genet.* 260, 574-581
- Gelvin, S.B. (2000) *Agrobacterium* and plant genes involved in T-DNA transfer and integration. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* 51, 223-256
- Gheysen G., Villarroel R., Van Montagu M. (1991) Illegitimate recombination in plants: a model for T-DNA integration. *Genes & Dev.* 5, 287-297
- Halfiter U., Morris P.-C., Willmitzer L (1992) Gene targeting in *Arabidopsis thaliana*. *Mol. Gen. Genet.* 231, 186-193
- Hanin M., Mengiste T., Bogucki A., Paszkowski J. (2000) Elevated levels of intrachromosomal homologous recombination in *Arabidopsis* overexpressing the MIM gene. *Plant J.* 24, 183-189
- Hansen G., Chilton M.-D. (1996) 'Agrolistic' transformation of plant cells: Integration of T-strands generated in planta. *Proc. Natl. Acad. Sci. USA* 93, 14978-14983
- Hansen G., Shillito R., Chilton M.-D. (1997) T-strand integration in maize protoplasts after codelivery of a T-DNA substrate and virulence genes. *Proc. Natl. Acad. Sci. USA* 94, 11726-11730
- Hohn B., Puchta H. (1999) Gene therapy in plants. *Proc. Natl. Acad. Sci. U.S.A.* 96, 8321-8323
- Kempin S. A., Liljegren S. J., Block L. M., Rounsley S. D., Yanofsky M. F., Lam E. (1997) Targeted disruption in *Arabidopsis*. *Nature (London)* 389, 802-803
- Kohli A., Gahakwa D., Vain P., Laurie D.A., Christou P. (1999) Transgene expression in rice engineered through particle bombardment: molecular factors controlling stable expression and transgene silencing. *Planta* 208:88-97
- Koncz K., Nemeth K., Redei G.P. and Schell J. (1994) Homology recognition during T-DNA integration into the plant genome, in J. Paszkowski (ed.) *Homologous recombination and gene silencing in plants*, Kluwer Academic Publishers, Dordrecht, the Netherlands, pp. 167-189
- Kooter J.M., Matzke M.A., Meyer P. (1999) Listening to the silent genes: transgene silencing, gene regulation and pathogen control. *Trends Plant Sci.* 4, 340-347
- Kumar S., Fladung M. (2000a) Transgene repeats in aspen: molecular characterization suggests simultaneous integration of independent T-DNAs into receptive hotspots in host genome. *Mol. Gen. Genet.* 264: 20-28
- Kumar S., Fladung M. (2000b) Determination of transgene repeat formation and promoter methylation in transgenic plants. *BioTechniques* 28:1128-1137
- Kumar S., Fladung M. (2001a) Gene stability in transgenic aspen (*Populus*). II. Molecular characterization of variable expression of transgene in wild and hybrid aspen. *Planta* DOI 10.1007/s004250100535.

- Kumar S., Fladung M. (2001b) Controlling transgene integration in plants. *Trends Plant Sci.* 6, 155-159
- Lebel E.G., Masson J., Bogucki A., Paszkowski J. (1993) Stress-Induced Intrachromosomal Recombination in Plant Somatic Cells. *Proc Natl Acad Sci U S A* 90, 422-426
- Lee K.Y., Lund P., Lowe K., Dunsmuir P. (1990) Homologous recombination in plant cells after *Agrobacterium*-mediated transformation. *Plant Cell* 2, 415-425
- Lloyd A.M., Davis R.W. (1994) Functional expression of the yeast FLP/*FRT* site-specific recombination system in *Nicotiana tabacum*. *Mol. Gen. Genet.* 242, 653-657
- Lyznik L.A., Mitchell J.C., Hirayama L., Hodges T. (1993) Activity of yeast FLP recombinase in maize and rice protoplasts. *Nucleic Acid Res.* 21, 969-975
- Lyznik L.A., Hirayama L., Rao K.V., Abad A. and Hodges T. (1995) Heat-inducible expression of *FLP* gene in maize cells. *Plant J.* 8, 177-186
- Maloisel L., Rossignol J.-L. (1998) Suppression of crossing-over by DNA methylation in *Ascombolus*. *Genes Dev.* 12, 1381-1389
- Matsumoto S., Ito Y., Hosoi T., Takahashi Y. and Machida Y. (1990) Integration of *Agrobacterium* T-DNA into a tobacco chromosome: possible involvement of DNA homology between T-DNA and plant DNA. *Mol. Gen. Genet.* 224, 309-316
- Mayerhofer R., Koncz-Kalman Z., Nawrath C., Bakkeren G., Crameri A., Angelis K., Redei G.P., Schell J., Hohn B., Koncz C. (1991) T-DNA integration: a mode of illegitimate recombination in plants. *EMBO J* 10, 697-704
- Mengiste, T., Revenkova, E., Bechtold, N., Paszkowski, J. (1999) An SMC-like protein is required for efficient homologous recombination in *Arabidopsis*. *EMBO J.* 18, 4505-4512
- Meyer P., Saedler H. (1996) Homology-dependent gene silencing in plants. *Annu. Rev. Plant Mol. Biol.* 47, 23-48
- Meyer P., Linn, F., Heidmann I., Meyer A, H., Niedenhof I., Saedler H. (1992) Endogenous and environmental factors influence 35S promoter methylation of a maize A1 gene construct in transgenic petunia and its colour phenotype. *Mol. Gen. Genet.* 231, 345-352
- Miao Z.H., Lam E. (1995) Targeted disruption of the TGA3 locus in *Arabidopsis thaliana*. *Plant J.* 7, 359-365
- Odell J.T., Russell S.H. (1994) Use of site-specific recombination systems in plants, in J. Paszkowski (ed.) *Homologous recombination and gene silencing in plants*, Kluwer Academic Publishers, Dordrecht, the Netherlands, pp. 219-270.
- Offringa R, de Groot M.J., Haagsman H.J., Does M.P., van den Elzen P.J., Hooykaas P.J. (1990) Extrachromosomal homologous recombination and gene targeting in plant cells after *Agrobacterium* mediated transformation. *EMBO J.* 9, 3077-3084
- Paszkowski J., Baur M., Bogucki A., Potrykus I. (1988) Gene targeting in plants. *EMBO J.* 7, 4021-4026
- Perrin A., Buckle M., Dujon B. (1993) Asymmetrical recognition and activity of the I-SceI endonuclease on its site and on intron-exon junctions. *EMBO J.* 12, 2939-2947.
- Puchta H. (1998) Towards targeted transformation in plants. *Trends Plant Sci.* 3, 77-78
- Puchta H. (2000) Removing selectable marker genes: taking the shortcut. *Trends Plant Sci.* 5, 273-274
- Puchta H., Dujon B, Hohn B. (1993) Homologous recombination in plant cells is enhanced by in vivo induction of double strand breaks into DNA by a site specific endonuclease. *Nucleic Acid Res.* 21, 5034-5040
- Puchta H., Swoboda P., Hohn B. (1995) Induction of intrachromosomal homologous recombination in whole plants. *Plant J.* 7, 203-210
- Puchta, H., Hohn, B. (1996) from centiMorgans to base pairs: homologous recombination in plants. *Trends Plant Sci.* 1, 340-348
- Puchta H., Dujon B., Hohn B. (1996) Two different but related mechanisms are used in plants for the repair of genomic double-strand breaks by homologous recombination. *Proc Natl Acad Sci U S A* 93:5055-5060

- Reiss, B., Klemm, M., Kosak, H., Schell, J. (1996) RecA protein stimulates homologous recombination in plants. *Proc. Natl. Acad. Sci. USA* 93, 3094-3098
- Reiss B., Schubert I., Köpchen K., Wendeler E., Schell J., Puchta H. (2000) RecA stimulates sister chromatid exchange and the fidelity of double-strand break repair, but not gene targeting, in plants transformed by *Agrobacterium*. *Proc. Natl. Acad. Sci. U.S.A.* 97, 3358-3363
- Reski R. (1998) *Physcomitrella* and *Arabidopsis*: the David and Goliath of reverse Genetics. *Trends Plant Sci.* 3, 209-210
- Risseuw E., Franke-van Dijk M.E., Hooykaas P.J. (1997) Gene targeting and instability of *Agrobacterium* T-DNA loci in the plant genome. *Plant J.* 11, 717-728
- Roberston K.D., Wolffe A.P. (2000) DNA methylation in health and disease. *Nat. Rev.* 1, 11-19
- Rong Y.S., Golic K.G. (2000) Gene targeting by homologous recombination in *Drosophila*. *Science* 288, 2013-2018
- Rouet P., Smih F., Jasin M. (1994) Expression of a site-specific endonuclease stimulates homologous recombination in mammalian cells. *Proc Natl Acad Sci U S A* 91, 6064-6068
- Shalev G., Sitrit Y., Avivi-Ragolski N., Lichtenstein C., Levy A. A. (1999) Stimulation of homologous recombination in plants by expression of the bacterial resolvase RuvC. *Proc. Natl. Acad. Sci. USA* 96, 7398-7402
- Spertini D., Beliveau C., Bellemare G. (1999) Screening of transgenic plants by amplification of unknown genomic DNA flanking T-DNA. *BioTechniques* 27:308-314.
- Srivastava V., Anderson O.D., Ow D.W. (1999) Single copy transgenic wheat generated through the resolution of complex integration patterns. *Proc. Natl. Acad. Sci. U.S.A.* 96, 11117-11121
- Stam M., Mol J.N.M., Kooter J.M. (1997) The silence of genes in transgenic plants. *Ann Bot* 79:3-12
- Sugita K., Kasahara T., Matsunaga E., Ebinuma H. (2000) A transformation vector for the production of marker-free transgenic plants containing a single copy transgene at high frequency. *Plant J.* 22, 461-469
- Thykjaer T., Finnemann J., Schauser L., Christensen L., Poulsen C., Stougaard, J. (1997) Gene targeting approaches using positive-negative selection and large flanking regions. *Plant. Mol. Biol.* 35, 523-530
- Tinland B., Hohn B. (1995) Recombination between prokaryotic and eukaryotic DNA: Integration of *Agrobacterium tumefaciens* T-DNA into the plant genome, in J.K. Setlow (ed.) *Genet Engineering Vol. 17*, Plenum press, New York pp. 209-229
- Yoon K., Cole-Strauss A., Kmiec E. B. (1996) Targeted gene correction of episomal DNA in mammalian cells mediated by a chimeric RNA·DNA oligonucleotide. *Proc. Natl. Acad. Sci. USA* 93, 2071-2076
- Zhu T., Peterson D. J., Tagliani L., St. Clair G., Baszczyński C., Bowen B. (1999) Targeted manipulation of maize genes in vivo using chimeric RNA/DNA oligonucleotides *Proc. Natl. Acad. Sci. USA* 96, 8768-8773
- Zubko E., Scutt C., Meyer, P. (2000) Intrachromosomal recombination between attP regions as a tool to remove selectable marker genes from tobacco transgenes. *Nat. Biotechnol.* 18, 442-445
- Zupan J., Muth T.R., Draper O., Zambryski P. (2000) The transfer of DNA from *agrobacterium tumefaciens* into plants: a feast of fundamental insights. *Plant J* 2000 23, 11-28

19

PLANT DNA METHYLATION AND GENE EXPRESSION

Michele Bellucci, Francesco Paolocci, Francesco Damiani, Sergio Arcioni

Istituto di Ricerche sul Miglioramento Genetico delle Piante Foraggere, CNR, Via della Madonna Alta, 130, 06128, Perugia, Italy.

INTRODUCTION

Although the necessity of interpreting the different behaviours of supposedly stable structures (the genes) and the visible differentiation of cells was outlined long ago (1934) by TH Morgan (reviewed in Jorgensen, 1994), little attention was paid to the reversible changes in the genes till the pioneering works of McClintock in the early 60s. More recently, even for the evidences coming out from the application of recombinant genetics, considerable attention has been devoted to epigenetics. Epigenetics is defined as “the study of mitotically and/or meiotically heritable changes in the function of a gene that cannot be explained by changes in its DNA sequence, and methylation is a DNA modification which has been postulated to play a central role in epigenetic regulation by modulating access to genetic information (Riggs, 1975; Holliday and Pugh, 1975).

Research on DNA methylation has been mainly focused on the biological significance of the addition of a methyl group to the cytosine ring, a DNA modification common in both prokaryotes and eukaryotes. Indeed, methylated cytosines are found in the genome of many organisms (Vanyushin et al., 1968; 1970; Shapiro, 1970), although some of the genetically tractable model organisms such as *Drosophila melanogaster*, *Caenorhabditis elegans* and *Saccharomyces cerevisiae* do not have methylated cytosines in their genome. In

prokaryotes, modification of DNA by methyltransferases regulates several cellular processes, including cleavage of DNA by the cognate restriction endonucleases, transposition, DNA repair, and transcription (Noyer-Weidner and Trautner, 1993). The covalent modification of DNA by cytosine methylation is involved in the regulation of a number of biological processes, which can all be explained as DNA-silencing events, in vertebrate animals, plants and fungi (for reviews see Finnegan et al., 1998; Colot and Rossignol, 1999). This modification can modulate gene expression by altering recognition of the double helix by the transcriptional machinery and the structural proteins that assemble chromatin (Nan et al., 1997; Kass et al., 1997). DNA methylation can control gene activity either at a small by affecting a single promoter and enhancer, or at large extent through global mechanisms that influence many genes within an entire chromosome or genome.

Cytosine methylation in plants is thought to play a role in the regulation of gene expression, in the timing of DNA replication, and in other aspects of plant development (Richards, 1997; Finnegan et al., 1998). In animal systems, DNA methylation is essential for development in the mouse and plays an important role in inactivating the X-chromosome and parental imprinting (Mohandas et al., 1981; Beard et al., 1995). It may also contribute to other activities, such as determination of the chromatin structure, tissue-specific gene transcription, and frequency of mutation, but conclusive evidence on these points is lacking (for reviews see Jost and Saluz, 1993; Bestor and Verdine, 1994; Razin, 1998; Bird and Wolffe, 1999). Cytosine methylation has also been shown to suppress homologous recombination (Maloisel and Rossignol, 1998). Moreover, several lines of evidence support the hypothesis that a primary function of CpG methylation in mammals and plants is to protect the genome from the expression and transposition of parasitic and exogenous DNA elements (Yoder et al., 1997; Goubely et al., 1999; Matzke and Matzke, 1998; Fagard and Vaucheret, 2000). Furthermore, increased or altered patterns of DNA methylation are one potential explanation for rapid genome change in nascent polyploids, and thus may be relevant to genome evolution in polyploid plants (Wendel, 2000).

The major theme of this chapter will be the role of DNA methylation in regulating plant gene expression.

NUCLEOTIDIC TARGET OF DNA METHYLATION

The DNA of higher eukaryotes is methylated at carbon 5 of some cytosine residues. The proportion of cytosines that are modified to 5-methylcytosines (m^5C) ranges from 3-8% in vertebrates (Shapiro, 1976) to more than 30% in several plant species (Shapiro, 1976; Matassi et al., 1992; Wagner and Capesius, 1981). In the vertebrate genome, 5-methylcytosine (5mC) is largely confined to dinucleotide CpG, whereas in higher plants 5mC is present in any sequence context of the nuclear genome (Meyer et al., 1994; Oakeley and Jost, 1996; Wang et al., 1996), although it is more frequent in dinucleotide CpG and trinucleotide CpNpG (Gruenbaum et al., 1981; McClelland, 1983; Jeddeloh and Richards, 1996). Eukaryotic genomes are not methylated uniformly, but contain specific methylated regions, with other domains remaining unmethylated (Bird, 1995). Plants, like animals, contain unmethylated CG rich regions, related to transcriptionally active genes such as the maize genes *A1* (dihydroflavonol reductase) and *Adh1* (alcohol dehydrogenase (Antequera and Bird, 1999; Nick et al., 1986). Antibodies to methylcytosine have shown that different regions of plant chromosomes have different levels of methylation (Frediani et al., 1996; Castilho et al., 1999).

The differences in 5mC content and distribution between vertebrates and plants might indicate distinct or multiple functions for DNA methylation in different eukaryotic systems. The higher content of methylated cytosine in some plants could be partly attributed to the large genome which contains many repetitive DNA sequences (Belanger and Hepburn, 1990), although an increase in plant genome size is not always correlated with an increase in the level of methylcytosine (Goubeley et al., 1999). Indeed, in plants methylation of cytosine residues is mainly restricted to the nuclear genome

(Finnegan et al., 1998), and is a common feature of all types of repeated sequences (Gardiner-Garden et al., 1992; Richards, 1997).

ENZYMES INVOLVED IN DNA METHYLATION

The structural and functional analysis of the key enzymes involved in DNA methylation is a critical step toward understanding how gene expression might be regulated by this epigenetic phenomenon. Cytosine 5-methyltransferase (Mtase) catalyses the transfer of methyl groups from the cofactor S-adenosyl methionine to the C5 position of cytosine in DNA. MTases in bacteria, fungi, plants and mammals all contain a methyltransferase domain with eight conserved motifs separated by more variable regions (reviewed in Finnegan et al., 1998). There are two major types of DNA MTase activities, maintenance and de novo. The methylation of hemi-methylated CpG and CpNpG symmetrical motifs after DNA replication is maintenance methylation. It has been proposed (semi-conservative model) that methylation of these symmetrical sequences provides a signal for transmitting methylation patterns through cycles of cell division based on the action of a maintenance MTase which methylates C residues in a newly synthesised strand, if the opposite strand carries a 5mC residue in the complementary sequence (Holliday and Pugh, 1975; Riggs, 1975). This results in stable patterns of methylation that are maintained throughout development or, in many cases, between generations. Cytosine methylation that occurs at previously unmethylated residues is known as de novo methylation. For symmetric CpG and CpNpG sequences, de novo methylation needs to occur only once, after which methylation can be preserved by maintenance activity. Methylation at non-symmetric sites was probably not efficiently transmitted to the newly synthesised daughter strand in the semi-conservative model. Cytosine methylation of non-symmetrical sequences was reported in mammals and fungi (Crowther et al., 1989; Selker et al., 1993), and could contribute greatly to the regulation of gene expression (Dieguez et al., 1997). Pelissier and

colleagues (1999) suggested that likely non-symmetrical methylation patterns have to be established de novo after each DNA replication cycle, and that it is not clear if the processes involved in de novo methylation of symmetrical sequences are different from those taking place in the de novo methylation of non-symmetrical sequences.

In plants, the characterized DNA MTases are closely related to the mammalian enzymes. Riggs (1975) and Holliday and Pugh (1975) predicted the existence of a maintenance MTase and the existence of such Mtase was subsequently demonstrated in mammalian cells (Wigler et al., 1981; Stein et al., 1982). The cDNA and gene, named *Dnmt1*, encoding this enzyme have been cloned from mouse (Bestor et al., 1988; Yoder et al., 1996) first and afterward on many other species.

Purified Dnmt1 protein methylates DNA containing hemi-methylated CpG dinucleotides more efficiently than unmethylated DNA in vitro (Bestor, 1992). Further, inactivation of the mouse *Dnmt1* gene by gene targeting results in extensive demethylation of all sequences examined (Li et al., 1992; Lei et al., 1996). Together, these findings strongly suggest that Dnmt1 functions as a major maintenance MTase in vivo. Recently, another abundant human species of CpG MTase enzyme, whose mRNA results from alternative splicing of the primary *Dnmt1* transcript, was described (Hsu et al., 1999). The originally described and the recently discovered forms of CpG MTase have been named Dnmt1-a and Dnmt1-b, respectively. Although Dnmt1-a can methylate unmethylated DNA in vitro and Dnmt1-a expressed in a baculovirus expression vector can methylate unmethylated DNA in vivo (Pradhan et al., 1999), evidence for its involvement in de novo methylation during development remains sketchy. Enzymes other than Dnmt1 might be responsible for de novo methylation in vivo. Several new species of MTases have been cloned from mouse and human and designated Dnmt2, Dnmt3-a and Dnmt3-b (Yoder and Bestor, 1998; Okano et al., 1998; Okano et al., 1999; Xie et al., 1999). In the mouse, Dnmt3-a and Dnmt3-b are two DNA MTases that catalyse the methylation of unmethylated and hemi-methylated substrates with equal efficiency (Okano et al., 1998). Inactivation of both genes by gene targeting blocks de novo

methylation in mouse embryonic stem cells and early embryos, but it has no effect on maintenance of the imprinted methylation pattern (Okano et al., 1999). Together these results indicate that Dnmt3-a and Dnmt3-b both function as de novo MTases, playing an important role in normal development and disease.

The most detailed studies on plant MTases have been conducted on the dicot species *Arabidopsis thaliana*. The first gene to be cloned was METI (Finnegan and Dennis, 1993), but an antisense construct of METI decreased the methylation of cytosines in CpG and CpCpG motifs, with little effect on the methylation of CpAp/TpG trinucleotides, suggesting that another enzyme methylates these sites (Finnegan et al., 1996). Later, five additional MTase genes were identified (Genger et al., 1999).

DNA MTase genes in this plant can be grouped in three classes (reviewed in Finnegan and Kovac, 2000), and this is probably typical of plants in general, as all families are represented in the monocot species *Zea mays* (Finnegan et al., 2000). The first class of MTases, which is formed by proteins similar to the mouse Dnmt1 MTase, include at least four enzymes from *A. thaliana* (METI, METII-a, METII-b and METIII), and also enzymes from other species: the two carrot MTase and the pea MTase (Pradhan et al., 1998; Bernacchia et al., 1998). METIII is likely not essential because it encodes a truncated protein, the METII genes are transcribed in most tissues even if at a low level. The presence of the METI antisense did not lower the level of METII-a and -b transcripts but reduced DNA methylation, indicating that the METII enzymes do not substitute for METI. The second class of MTases, which appear to be unique to plants, is formed by the *A. thaliana* CMTI and CMTII enzymes (Genger et al., 1999), encoding proteins having both a chromodomain, a short motif found in chromatin-associated proteins inserted between motifs II and IV (Henikoff and Comai, 1998). The third class of *A. thaliana* putative MTases is most similar to the mammalian Dnmt3 family of de novo MTases, suggesting that they encode plant de novo MTases (Okano et al., 1998).

Which are the possible roles in methylating DNA of these several MTases in *A. thaliana*? It is probably too early to answer this question

because target specificity (CpG and CpCpG) has been determined only for METI (Finnegan 1996), while indispensable data on the role of individual MTases in the establishment or maintenance of methylation patterns are still lacking, as are analyses of knockout mutants or antisense transformants for each gene. However, METI is likely to be the predominant maintenance MTase and the CMTI family of MTases can be associated with heterochromatin (Paro and Harte 1996). In *Pisum sativum*, the characterisation and in vitro expression of a cDNA encoding a 5mC MTase (~174 kDa) suggest that the target sequence of the enzyme were the dinucleotide CpG and trinucleotides CpAp/TpG (Pradhan et al., 1998). Being rich in basic amino acid, the pea enzyme could be proteolytically processed to smaller species (100, 110, and 140 kDa) with a different target specificity keeping the MTase domain intact, as previously observed (Pradhan and Adams, 1995). The presence of smaller proteins in wheat and rice DNA MTase enzyme preparations also supports the idea that proteolysis plays a role in enzyme maturation (Theiss et al., 1987; Giordano et al., 1991).

Up to now, there is only evidence supporting the notions that i) plant MTases may differ in sequence specificity; ii) generally MTases are more active in meristematic cells than in other tissues (Ronemus et al., 1996; Bernacchia et al., 1998). The most significant difference between plant and animal Mtases is a 40-41 amino acid deletion in the plant MTase variable region between conserved motifs VIII and IX, referred to as the target recognition domain (TRD), which determines the sequence specificity of methylation in prokaryotic C-5 MTase. This may reflect differences in the interaction of plant and animal Mtases with the DNA template.

Recently, a mammalian protein with specific demethylation activity for methyl CpG dinucleotides was described, together with a demethylase enzyme complex that acts processively converting 5-methylcytosine to cytosine and methanol (for review see Wolffe et al., 1999). This active demethylation process is presumably also present in plants, as suggested by changes in DNA methylation during tobacco pollen development (Oakeley et al., 1997).

METHODS TO REVEAL DNA METHYLATION

The methods for assessing DNA methylation can be divided into two categories: 1) direct, or structural and 2) indirect, or functional.

The first consist in the analysis of DNA by classical methodologies of biochemistry such as spectrophotometric and HPLC analyses. The first studies conducted in this field (Vanyushin 1968) utilised paper chromatography to fractionate methylated nucleotides and spectrophotometric measurements for quantification. An improvement of these techniques was the use of HPLC (Wagner and Capesius, 1981). All these methods had a quantitative relevance because they allowed determining the percentage of methylated nucleotides and therefore verifying if a given treatment can increase or decrease the level of methylation. The antibody technique opened new opportunities in this area. Oakeley and co-authors (1997) utilised an anti-5mC antibody to measure the decrease in DNA methylation occurring in the generative nucleus compared to the vegetative ones in the pollen cells of tobacco.

The methods described previously are useless for a qualitative evaluation of methylation, specifically to determine which sequences are methylated and where methylation is located. The different sensitivity to methylation of restriction enzymes provided the tool for analysing site specific methylation. In fact, some enzymes are unable to cut when C is methylated in their recognition sequence, therefore restriction with such enzymes of known sequences would result in the absence of the expected DNA fragment if methylation occurred at the cleavage site (Cedar et al., 1979). The more suitable enzymes are those which have isoschizomers with a different sensitivity to methylation, which is why the enzymes *HpaII* and *MspI* (recognition sequence: CCGG) are the most frequently utilised for this purpose. *HpaII* does not cut the DNA when either of the C is methylated. *MspI* does not cut when the 5' C is methylated but, is active when the 3' C is methylated. Methylation of this latter C discriminates the restriction of the two enzymes and consequently can be detected. To test the efficiency of restriction and hence the methylation of target sequences,

DNA can be submitted to Southern analysis or to PCR amplification utilising primers complementary to flanking regions of the restriction sites. The main limiting factor of these techniques is the number of restriction sites of methylation sensitive enzymes in the target sequences. A recent improvement of this technique consists on a first DNA restriction necessary for producing landmark fragments, which are made radioactive in the terminal sequences, electrophoresed in capillary agarose gel, restricted with a second enzyme when still embedded in the capillary gel, then fused with 5% polyacrilamide to separate DNA in a second dimension gel (Okazaky et al., 1997) This method is called restriction landmark genome scanning (RLGS) and is not specific for studies concerned with DNA methylation, but can be profitably used for this purpose using restriction enzymes sensitive to methylation and may offer the possibility to visualise the level of methylation of an entire genome. With this approach the role of methylation in determining altered phenotypes in *A.thaliana ddm1* mutants is analysed ([http://ss.abr.affrc.go.jp/organization/Molecular Genetics](http://ss.abr.affrc.go.jp/organization/Molecular%20Genetics), web page of the Laboratory of genome function, Japanese National Institute of Agrobiological Resources).

An other method has been developed, which is independent from restriction (Frommer et al., 1992). It is based on the observation that when single stranded DNA is treated with sodium bisulfite more than 96% of cytosine converts to uracil against a mere 2-3% for 5mC, which therefore remains largely unreactive (Wang et al., 1980). The method consists of the following steps: denaturing restricted or sheared DNA; treating with sodium bisulfite; PCR amplifying of the two strands with two primers per strand designed on the basis of the modified sequence, sequencing of amplified products. In this way it is possible to identify a single methylated cytosine. Independently from the availability of restriction site, with this method Ronchi et al. (1995) detected the methylation dependent inactivation of repeated endogenous genes in maize.

The indirect or functional approach to test the relation between gene expression and methylation consists in applying demethylating agents to individuals carrying the silenced target gene and then testing for the recovery of gene expression. The most common demethylating

chemical is azacytidine (AZA) (Jones and Taylor, 1980), but ethionine (Cox and Irving, 1977) and amino-ethoxy-carbonil-pyrimidine (ECP) (Raugei et al., 1981) are also used to demethylate DNA sequences. The main aspect to keep in consideration is the toxicity of these compounds because it is quite obvious that an extensive demethylation of the genome could result in the complete deregulation of cell activity.

MUTANTS FOR UNDERSTANDING THE ROLE OF DNA METHYLATION

The most intriguing yet unanswered question is why and how DNA is methylated. There are no final answers, however individuals mutated for their ability in establishing standard levels of DNA methylation have proved useful tools for working in this field. Using these mutants to study the functions of DNA modification in eukaryotes is an attractive approach that avoids some limitations associated with correlative studies and the use of methylation inhibitors (Kakutani et al., 1999). Eukaryotic mutants affecting genomic DNA methylation have been described in many organisms.

In plants, DNA hypomethylation mutants have been isolated in *A.thaliana*: the *ddm1* mutants (for decrease in DNA methylation) identified by screening mutagenised populations for plants containing a centromeric repetitive DNA array susceptible to digestion by endonuclease sensitive to methylated cytosine (Vongs et al., 1993); and *met1* mutants, expressing *MET1* in an antisense orientation (Finnegan et al., 1996; Ronemus et al., 1996).

The *ddm1* mutants showed a reduction of over 70% of 5-methylcytosine, with hypomethylation mainly occurring on a large variety of repeated DNA families (Vongs et al., 1993). Despite the fact that homozygous *ddm1* mutants initially displayed only weak morphological changes, a number of aberrant phenotypes were observed through propagation of the *ddm1* mutations by repeated self pollination (Kakutani et al., 1996). After six generations of selfing, the *ddm1* plants showed a reduction or an increase in apical dominance,

short internode length, late flowering, small leaf size, reduced fertility, and homeotic changes of flowers. Selfing produced a progressive and low reduction of cytosine methylation also in non-telomeric single-copy sequences. *ddm1* mutations lead to the formation of heritable lesions at unlinked loci that cause the morphological phenotypes. These lesions were stably transmitted, even when segregated apart from the *ddm1* mutation. Such a mutation is completely recessive, and re-methylation of DNA was not observed during repeated backcrosses of a *ddm1* mutant to wild type plants (Kakutani et al., 1999). Recently, it has been reported that *DDM1* encodes a SW12/SNF2-like protein (Jeddeloh et al., 1999). These proteins are reported to be implied in nucleosome-remodelling complexes, transcriptional activation and repression, chromatin assembly and DNA repair (Sudarsanam and Winston, 2000). Another protein, MOM1, has been recently identified, at least partially, as a member of SW12/SNF2-like protein (Amedeo et al., 2000). However, *mom1* mutants, even after nine generations of selfing, do show neither phenotypic abnormalities nor DNA hypomethylation, implicating that MOM1 affects fewer loci than does DDM1 and/or acts downstream of DDM1.

met1 mutants show a reduction of up to 90% in cytosine methylation at the centromeric DNA and rDNA. Differently from *ddm1* which affects methylation of both CpG and CpNpG, *met1* preferentially reduce methylation in CpG dinucleotides (Finnegan et al., 1996). Unlike *ddm1*, *met1* display a substantial demethylation of both repetitive and single-copy gene sequences Ronemus et al. (1996). The spectrum of phenotypes in these lines cover most of the developmental abnormalities reported in the *ddm1* including alteration of flowering time, decreased fertility, alteration of leaf size and shape, changes in meristem identity and organ number, reduced apical dominance (Ronemus et al., 1996; Finnegan et al., 1996). The phenotypes became more severe in successive generations of *met1* lines even though the level of methylation was unaffected and, similarly to *ddm1*, re-methylation of DNA occurred slowly in antisense-null progeny of a hemizygous *met1* plant (Finnegan et al., 1996). It is noteworthy that, despite a general demethylation effect on the entire genome, both *ddm1* and *met1* cause an unexpected

hypermethylation of previously hypomethylated alleles of the floral developmental *SUPERMAN* (*SUP*) and *AGAMOUS* (*AG*) genes, and methylation was associated with transcriptional repression of these genes (Jacobsen and Meyerowitz, 1997; Jacobsen et al., 2000). This observation has also been reported in mammals, where in the early stage of human carcinogenesis, silencing and ectopic methylation of tumour suppressor genes occur amid a background of DNA hypomethylation (Jones and Laird, 1999). The mechanism leading to this phenomenon is still not clear. A possible cause could be the induction of a compensatory methylation activity with altered specificity or the generation of regions of chromatin more accessible to methylation (Stoke and Richards, 2000).

An evident role of DNA methylation controlling gene expression has been reported by Finnegan et al. (1998) who showed that flowering promotion is directly proportional to the decrease in DNA methylation on *met1* lines of late flowering *A. thaliana* ecotypes, which require a prolonged period at low temperature (vernalization) before they will initiate flowering. In addition, both *ddm1* and *met1*, as well as treatment of plants with AZA resulted in early flowering of such ecotypes, suggesting that vernalization could be mediated by demethylation of promoters of genes implied in flowering. The *FLC* (for flowering locus C) gene coding for a MADS-box protein has been recently characterised as a repressor of flowering in *A.thaliana* (Sheldon et al., 2000), and its transcription is dramatically reduced both in plants showing reduction of DNA methylation and in plants after vernalization. Since a decrease in methylation generally enhanced transcription, it is surprising that demethylation in this case results in a reduction of gene transcription, and it is still unknown if methylation regulates *FLC* expression directly or indirectly by controlling the transcription of a regulator of *FLC* (Finnegan et al., 2000).

DNA METHYLATION AND REPETITIVE ELEMENTS

DNA methylation has been proposed to operate as a defence system against parasitic mobile elements, in order to save the integrity

of genomes (Matzke et al., 1996). Transposons may induce detrimental changes in the host genomes in several ways, one of which is the disruption of genes by insertion mutations. In plants, as much as 40 to 80% of the genome can be occupied by retrotransposons (Heslop-Harrison, 2000). However, only a small portion of spontaneous mutations has been shown to be caused by retrotransposons (Wessler et al., 1995), which are highly methylated and almost always not transcribed (Bennetzen et al., 1994; Hirochika H, 1997). Conversely, in *D.melanogaster*, in which DNA methylation does not occur, retrotransposons caused 80% of spontaneous mutations (Green, 1988). Therefore, it is conceivable that methylation is an effective strategy in retrotransposon suppression. Recently, Hirochika and colleagues (Hirochika et al., 2000) suggested that the inactivation of retrotransposons and the silencing of repeated genes have mechanisms in common, showing that the introduction at high copy number of tobacco *Tto1* retrotransposon (Hirochika et al., 1993) correlates with its DNA methylation and inactivation in *A.thaliana* wild type. On the contrary, in homozygous *ddm1*, the *Tto1* became hypomethylated, transcriptionally and transpositionally active. Interestingly, a silent endogenous *Tar17* retrotransposon was transcriptionally reactivated in these mutants, even if new transposition events were not detected. In addition, in *mom1*, two transcriptionally reactivated transcripts TSI-A and TSI-B have been identified. TSI are synthesised from repeats located in the heterochromatic pericentromeric regions of all five *A.thaliana* chromosomes, and share sequence homology with the 3' termini of retrotransposon *Athila*. In this case, again, no new transposition events have been detected (Steimer et al., 2000). Such transcripts have also been reactivated in most of the transcriptional gene silencing impaired mutants, reinforcing the idea that heterochromatin retroelements are endogenous silencing targets and that transcriptional silencing of repeated genes and retrotransposons may share a similar mechanism.

In mammals, the transcriptional control of *Alu* (SINEs) and *L1* (LINEs) retrotransposons is also controlled by methylation. DNA

methylation has been shown to repress the transcription of *Alu* sequences both in vitro and in vivo (Liu, 1993; Liu et al., 1994). It has also been suggested that methylation of repetitive elements can repress their expression by methylation-mediated transcriptional inactivation and by increasing the rates of C-T mutations over the course of evolution (Bestor, 1998).

In *Brassica napus*, the Sl_{Bn} (SINE) retroelements was shown to be highly methylated, both in symmetrical and asymmetrical positions, not only in transcriptionally important regions but also distributed along the entire sequence (Goubely, 1999). Sl_{Bn} methylation was more than twice the general methylation level in the host genome. Moreover the high level of its methylation is not related to any particular genomic localisation, suggesting that these elements were specifically targeted for methylation. *Sl* element methylation can also be spread into flanking genomic sequences thus creating a distal epigenetic modification (Arnaud et al., 2000), as it was also observed in maize for the *P* gene hosting an inactive *Ac* transposable element (Brutnell and Dellaporta, 1994).

The mobility of transposable elements such as maize *Ac*, *Mu* and *Spm* also is controlled by methylation (Brutnell and Dellaporta, 1994; Schlappi 1994), where somatic inactivation and reactivation is associated with methylation changes: low levels of transcription and mobility are correlated with extensive methylation. However, the activity of maize transposable elements, *Ac* and *Spm*, differed depending on whether they were inherited from a male or a female gamete (Fedoroff and Banks, 1988), implying that methylation can change during gametogenesis. *Spm* and *Mu* elements are also more heavily methylated in older leaves than in younger ones, a sure sign that DNA methylation of these elements increases with development (Banks et al., 1988; Martienssen 1994, cited by Finnegan 1998), and their demethylation probably occurred during gametogenesis or at an early stage of seed development (Finnegan et al., 1998). The increasing methylation of transposable elements and other repeated sequences during the development may then ensure that these sequences are packaged into heterochromatin before gamete formation. This in order to reduce meiotic recombination between

repeated sequences at different loci without causing major rearrangements leading to loss of gamete viability (Hsieh, 1992; Finnegan et al., 1998).

DNA METHYLATION AND GENE FAMILY EXPRESSION

Endogenous genes sharing a large degree of sequence homology (> 90%) were proved to undergo gene silencing (Ronchi et al., 1995). REED, for Reduction of Expression of Endogenous Duplicates is a phenomenon similar to paramutation described to occur in maize. Ronchi and colleagues showed that as a result of coupling *R* and *Sn* genes, two regulatory factors of the anthocyanin pigmentation, the *Sn* allele is partially silenced, leading to weakly pigmented seedlings. The decrease in tissue pigmentation, as revealed specifically in genomes hosting both *R* and *Sn* genes, was correlated with extensive methylation of the *Sn* promoter, not restricted to symmetrical CpG and CpNpG sequences and not associated to gross structure DNA rearrangements. Promoter demethylation and *Sn* transcriptional activity as well as tissue pigmentation were restored by AZA treatment, suggesting a role of DNA methylation in REED, but whether this was a consequence or a cause has not been established.

The *PAI* tryptophan biosynthetic gene family, coding for phosphoribosylanthranilate isomerase in *A.thaliana* also displays intriguing features. There is an unusually high degree of sequence identity among *PAI* family members, including promoters and introns (Melquist et al., 1999). The Wassilewskja (WS) genotype carries four *PAI* genes at three unlinked sites: a tail-to-tail inverted repeat at one locus (*PAI-PAI4*) plus singlet genes *PAI2* and *PAI3* at two other loci (Bender and Fink, 1995), and all four genes are densely methylated in regions of genes that have sequence identity to each other, but methylation does not spread more than a few hundred bases beyond the boundaries of *PAI* sequence identity (Luff et al., 1999). Differently, the Columbia (*Col*) genotype has three singlet *PAI* genes and no cytosine methylation. Methylation in WS is correlated with silencing of, at least, the *PAI2* gene. However, the expression of all *PAI* genes in this genotype, mainly due to the expression of *PAI1*, is

sufficient for a normal plant phenotype (Bender and Fink, 1995; Jeddeloh et al., 1998). It is also worth noting that in the WS strain, *PAI1* and *PAI2* genes, although displaying no significant differences in methylation, are differently expressed: the former being active and the latter silenced, suggesting a role of *cis*-acting sequences, DNA structures and /or chromosomal domains in determining a different chromatin configuration and expression states. *PAI* rearrangements, specifically the presence of methylated inverted repeats, has been correlated with *de novo* cytosine methylation both at symmetrical and asymmetrical sites of sister gene sequences elsewhere in the genome. In fact, the introduction into the *Col* genome of either a *PAI-PAI4* inverted repeat or a transgene consisting of a promoterless *PAI-PAI4* inverted repeat, triggers *de novo* cytosine methylation of the three unlinked *Col PAI* genes (Luff et al., 1999). Since no transgene RNA was detected, a DNA/DNA pairing mechanism has been postulated to trigger *de novo* methylation. Furthermore, the observation that *PAI1-PAI4* inverted repeats can methylate the allelic identical *PAI1* gene more rapidly than *PAI2* or *PAI3* genes suggested that during the DNA/DNA homology search, allelic sequences have a higher probability of pairing than the same sequence on a different chromosome position (*PAI2*) or with a slightly different sequence (*PAI3*). Finally, as with some *A.thaliana* transgenic lines showing a reactivation of a previously silenced transgene upon crosses with the *ddm1* mutant line, Jeddeloh et al. (1998) reported the demethylation and reactivation of the *PAI2* gene in the *ddm1* background.

Repeated sequences arranged as inverted repeats seem to be especially potent *trans*-acting inducers of gene silencing and methylation in transgenic lines (Stam et al., 1998; Mette et al., 1999, Waterhouse et al., 1998). The frequent silencing encountered in multicopy transformants has led to speculation that enhanced DNA-DNA pairing of the repetitive elements, even in an ectopic configuration, might act as a signal for their detection, resulting in highly efficient silencing. Evidence that repeat-induced methylation events are triggered by DNA/DNA pairing has been reported in fungi. Repeated-Induced Point mutations (RIP) in *Neurospora crassa* (Rossignol and Faugeron, 1994) and Methylation Induced

Premeiotically (MIP) in *Ascobolus immersus* (Selker 1997), are two examples of inactivation of repeated sequences by methylation that causes transcription arrest. In plants however recent data have demonstrated that inverted repeats-induced de novo methylation was based on RNA-directed DNA methylation, and it is likely that some observations of de novo methylation that were proposed to be DNA-mediated are indeed RNA-directed (for review see Wassenegger, 2000).

DNA METHYLATION AND GENETIC TRANSFORMATION

In the 1980s papers started to appear, which described atypical phenomena associated with the genetic transformation of plants. Specifically, overexpression of a chimeric transgene could result in “silencing”, inactivation of the introduced transgene, or the suppression of both transgene and homologous endogenous gene copies, the latter phenomenon being called “cosuppression” (Napoli et al., 1990; van der Krol, et al., 1990). Originally perceived as an unwelcome surprise, and an unwanted problem for companies aiming to produce genetically modified organisms (GMOs) with a stable integration and expression of the introduced genes, the phenomenon of silencing has supplied scientists with a new tool to discover the epigenetic mechanisms that regulate gene expression both in the nucleus and in the cytoplasm (Matzke and Matzke, 1995). In fact, it became rapidly clear that silencing was a general and frequent event, not limited to a particular chimeric gene or plant species, but rather a phenomenon occurring also in different eukaryotic systems such as fungi, ciliates, nematodes, fruit flies etc. (Cogoni and Macino 1999a, Ruiz et al., 1998; Fire et al., 1998; Kennerdell and Carthew, 1998). Thus, gene silencing is part of a universal gene regulation system. Two different classes of silencing events can be distinguished.

One class includes the events resulting from “position effect”, that is the negative influence exerted on transgene expression by flanking plant DNA and/or chromosomal location. This type of silencing therefore reflects the epigenetic state of neighbouring host sequences or the relative tolerance of a chromosome region to invasion by

exogenous DNA (Matzke and Matzke, 1998). However, in transgenic petunia (Prols and Meyer, 1992), single copy transgenes can be unstable expressed even when they are located in a nonmethylated area of the host genome. Factors such as the presence of backbone plasmid DNA transferred with the transgene (Iglesias et al., 1997), or the presence of repetitive elements representing hot spots for *de novo* methylation in transgenic constructs carrying reporter genes (ten Louis et al., 1995), can lead to inactivation, or induce a variegation of expression, of the transgenes. Moreover, a different GpC content between the transgene and the surrounding genomic sequences (Meyer et al., 1994; Elomaa, 1995) was found to be responsible for gene inactivation. Meyer and Heidmann (1994), reported that in the progeny of transgenic petunia transformed with the maize *Al* gene, the transgene, even inserted in a non methylated environment, was methylated, while its chromatin environment remained hypomethylated, and transgene specific methylation was not limited to the 35S promoter but also occurred within the coding region. The fact that the *Al* transgene differs significantly in its AT content from the chromosomal environment, and the inserted sequences are unstable in non-matching host genomic sequences, has suggested the presence of a genomic system, acting as a defence mechanism and capable of identifying and inactivating an “intruder” sequence through *de novo* methylation.

A second class is based on the presence of multiple copies of a particular sequence, and is designated as homology-dependent gene silencing (HDGS) (Meyer and Saedler, 1996). A further distinction can be made in HDGS between transcriptional gene silencing and post-transcriptional gene silencing events, TGS and PTGS, respectively. TGS reduces gene expression by suppressing transcriptional initiation, TGS can act in a *cis* arrangement, in which homologous regions, with closely linked or tandemly arranged, are silenced as reported in *A.thaliana* as repeat-induced gene silencing (RIGS), (Assaad et al., 1993), where multiple copies of a transgene lead to increased DNA methylation and changing chromatin configuration, and silencing is associated with the absence of steady-state mRNA (Yeand Signer, 1996). *Cis* acting elements may also be

endogenous heterochromatin surrounding the transgene locus (position effect). TGS also acts in a *trans* configuration, in which case an unidirectional effect of an inactivated transgene is transmitted to a target gene that may be in either an allelic or a non allelic chromosomal position, repressing its expression (Matzke et al., 1994; Vaucheret et al., 1993; Bellucci et al., 1999). This epigenetic phenomenon resembles “paramutation” (Brink, 1960), in fact, a *trans* acting element (a paramutagenic allele) has the potentiality to transfer its epigenetic state to, and weaken the activity of, a sensitive (paramutable) allele by direct DNA-DNA pairing, which leads to a meiotically heritable reduction in the transcription of the target locus (Matzke and Matzke, 1998.).

TGS requires interacting genes sharing homology in the promoter regions, where as little as 90 bp of homology between a silencing locus and the promoter of a target transgene is sufficient for triggering silencing and methylation (Vaucheret, 1993). Diéguez et al. (1998) showed that methylation at CpG and CpNpG sites in the promoter region is not a prerequisite for the initiation of epigenetic gene inactivation in the progeny of transgenic tobacco lines derived by crossing a line carrying a selectable marker gene driven by a 35S promoter devoid of CpG and CpNpG methylation acceptor sites, with the silencing tobacco line 271 containing inactivated and methylated copies of the 35S promoter. In fact, the depletion of methylation of CpG and CpNpG acceptor sites seems to redirect MTase to cytosine residues in other sequence contexts, resulting in their hypermethylation. Nevertheless, silencing not supported by methylation at symmetrical sites was released immediately in the absence of the silencer, while in an analogous experimental system, it has been reported to be persistent for transgene driven by the wild-type 35S promoter containing CpG or CpNpG sequences (Park et al., 1996). Therefore, Diéguez and colleagues suggested that methylation at conventional sites is secondary to the initiation of transgene silencing, but it is important for the maintenance of the silent state. In rice, the R1 progeny of transcriptionally silenced line JKA 52 carrying the *bar* gene as selectable marker showed transgene methylation and herbicide sensitivity; however, a large proportion (up to 70%) of

seedlings obtained by selfing the silenced R1 lines and germinated in the presence of 5-azacytidine, were herbicide-resistant. Reactivation of the transgene correlated with loss of methylation of the promoter *Ubi 1* driving the bar gene. Such a reactivation was however only transient since the silenced states was re-established due to the renewed methylation of the transgene sequence (Kumpatla and Hall, 1998).

PTGS involves sequence-specific RNA turnover in the cytoplasm, so that RNA is actively produced in the nucleus but failed to accumulate in the cytoplasm because of a degradative process related to RNA interference (RNAi) in animals and quelling in fungi (Fire, 1999; Cogoni and Macino 1999a). Unlike TGS, PTGS operates at the level of sequence-specific RNA degradation and acts against transgenes, endogenous genes and viral genes. PTGS of nuclear genes is a manifestation of a natural defence mechanism in plants that is induced by a wide range of viruses (Ratcliff et al., 1999). As a matter of the fact, RNA viruses replicating cytoplasmically can trigger PTGS when they carry elements from exons of host genes, a phenomenon referred to virus-induced gene silencing (VIGS) (Baulcombe, 1999). Alternatively, transgenic plants expressing a cDNA fragment derived from the genome of an RNA virus, display resistance towards viruses that are similar in nucleotide sequence (Baulcombe 1996), by a sequence-specific degradation of the viral RNA leading to the elimination of the viruses, a phenomenon called “recovery” (Lindbo et al., 1993). PTGS requires homology in the transcribed regions, and this is a highly specific process, since PTGS only silences genes or viruses with a high degree of sequence homology (>75%) to the transgene (Waterhouse et al., 1999). These regions are also frequently reported to be methylated, where methylation may also occur at non symmetrical sites of the DNA (English al. 1996; Ingelbrecht et al., 1994; Jones et al., 1998). However, PTGS can be triggered even in the absence of DNA methylation, examples are reported in *D. melanogaster* and *C. elegans* which are devoid of DNA methylation, or in *N. crassa* strain *dim2* defective for C methylation (Cogoni et al., 1996). Two factors have been shown to trigger PTGS: a high transcription rate of the transgene, and arrangements of the transgenic

locus in inverted repeats (Elamayan and Vaucheret, 1996; Que et al., 1997, Stam et al., 1998; Waterhouse et al., 1998). Models that implicate either aberrant RNA molecules (Baulcombe and English, 1996; Wasseneger and Pellisier, 1998) or double-stranded RNA (Metzlaff et al., 1997, Montgomery and Fire, 1998) as causes of PTGS have been proposed. All the models require the presence of duplex RNA either as product of direct synthesis of sense and antisense RNA through the inverted repeats, or by pairing between highly produced sense RNAs.

A deeper insight into the role of methylation and chromatin structure in TGS and PTGS could be gained by the identification of mutants lines affecting silencing. Following mutagenesis of *A. thaliana* lines showing TGS or PTGS of transgenes such as *hpt*, *uidA*, *CHS*, and selection for revertants showing their re-activation, mutants impairing either in TGS or in PTGS have been isolated. Five (HOG1, SIL1, SIL2, SOM/DDM1 and MOM1) different trans acting modifiers of TGS have been so far identified. Among the TGS modifiers, the *hog1* and *som/ddm1* mutations released silencing and showed reduction of DNA methylation at transgene loci and at centromeric repeats (Furner et al., 1998; Mittelsten Scheid et al., 1998). Therefore, these studies have established a link between silencing and methylation. Conversely, other TGS modifiers, such as *sil1*, *sil2* and *mom1*, although showing a reactivation of transgene(s), do not alter its (their) methylation state (Mittelsten Scheid et al., 1998; Amedeo et al., 2000). Interestingly, crosses between a line carrying a silent and hypermethylated *hpt* transgene with the *met1* line, produce progeny showing a hypomethylated yet silenced transgene (Mittelsten Scheid et al., 1998). Thus, these studies show that silencing and methylation are not inextricably linked

The PTGS modifiers have been characterized only recently. *sgs1*, *sgs2* and *sgs3* (for suppressor of gene silencing), have been identified on the mutagenised 35S-*uidA* transgene by their opposite effect respect to the *egs* (enhanced gene silencing) loci (Dehio and Shell, 1994). They govern specific features of PTGS since, unlike *som/ddm1* mutants, they do not release the TGS of the 35S-*hpt* transgene, but the PTGS of the 35S *Nia2* transgene (Elmayan et al., 1998; Mourrain et

al., 2000). All these mutants do not affect centromeric repeats methylation, and therefore can not be considered methylation mutants, though they modified transgene methylation. *sgs2* and *sgs3* showed no methylation of the reporter gene at CpNpG sites and a reduction of methylation at CpG sites, while *sgs1* showed a reduction of *uidA* methylation at CpNpG sites and no reduction at CpG sites. SDE-1 is a further *A. thaliana* locus required for PTGS (Dalmay et al., 2000). While *SGS3* does not show much similarity to any other known or putative protein, the *SGS2* and SDE-1 loci code for a protein homologous to the RNA-directed RNA polymerase firstly isolated in tomato (Schiebel et al., 1998) and later on *N.crassa* and *C.elegans*, QDE-1 and EGO-1, respectively (Cogoni and Macino, 1999b; Smardon et al., 2000) reinforcing the hypothesis that PTGS, quelling and RNAi are mechanistically related.

The role of DNA methylation on TGS and PTGS, and whether TGS and PTGS occur by the same mechanism is not yet clear (Grant, 1999), but it has been proposed that TGS may have mechanistic similarities to PTGS (Wasseneger et al., 1994), as the two processes could have in common a modification of the chromatin configuration and/or DNA methylation, as a result of specialised RNA products (i.e aberrant RNA). According to Mette et al. (1999), the origin and the localisation of aberrant transcripts may actually give rise to TGS or PTGS events. Nuclear aberrant transcripts originating from promoter sequences can trigger TGS and promoter methylation that is also meiotically heritable, while aberrant transcripts originating from the coding regions may cause methylation of these sequences and induce cycles of cytoplasmic RNA turnover that is reset at each sexual generation. The identification of a transgene locus that undergoes TGS and can trigger both TGS of promoter-homologous transgenes and PTGS of coding sequence-homologous target transgenes (Thierry and Vaucheret 1996, Fagard and Vaucheret, 2000) is an example showing that TGS and PTGS may share common steps. This is supported by the observation that *de novo* methylation occurred as a consequence of the presence of homologous RNA in PTGS induced by a cytoplasmically replicating plant virus (Jones et al., 1998), and a sequence-specific RNA directed *de novo* methylation of homologous

transgene has been reported following viroid replication in the nucleus of transgenic plants where cDNA copies of viroid were stably integrated (Wasseneger et al., 1994).

IS METHYLATION THE 'IMPRINT MARK' IN PLANTS?

Genomic imprinting refers to the silencing of one allele of a gene according to its parental origin (Barlow, 1995). Thus, the parental alleles must be distinguished or marked. Multiple elements have been proposed to convey such an allelic mark, and DNA methylation has emerged as an essential component of the imprinting mechanism (Brannan and Bartolomei, 1999). In mammals either the paternally-derived or the maternally-derived allele in the embryo and its derivatives can be silenced (Isles and Wilkinson, 2000), and the importance of DNA methylation in the mouse for the maintenance of imprinting has been widely demonstrated (Li et al., 1993; Lei et al., 1996). In plants, the first evidence of parental imprinting was discovered in maize endosperm (Kermicle and Alleman, 1990). All the genes that have been found to be imprinted in maize have a maternally inherited allele which is either undermethylated, or over-expressed or both, whereas the paternally inherited allele is more methylated or has a reduced level of expression (Martienssen, 1998). In maize endosperm, several genes show differential expression of the maternal and paternal alleles. They include zein genes (Lund et al., 1995a), α -tubulin genes (Lund et al., 1995b), *dze* gene (Chaudhuri and Messing, 1994), and the *r* gene (Kermicle and Alleman, 1990). Differential methylation of maternal and paternal zein alleles encoding 19 kDa and 22 kDa proteins was observed, with the paternal (repressed) allele remaining methylated whereas the maternal allele was demethylated (Lund et al., 1995a). Recent studies on seed development in *A. thaliana* emphasised the importance in plants of maternal contributions not only to endosperm but also to embryo development, and provided evidences for the existence in plants of paternal silencing during embryogenesis (reviewed in Russinova and

de Vries, 2000; Finnegan et al., 2000). In this species, the genes *FIS1/MEA*, *FIS2* and *FIS3* are expressed only by the maternally inherited allele (Finnegan et al., 2000 and references herein). *FIS* genes are involved in the suppression of seed development, including the embryo and the endosperm, in the absence of fertilisation, and after fertilisation, play a role in promoting embryo development (Luo et al., 1999). Vielle-Calzada and colleagues (2000) showed that none of the paternally inherited alleles of twenty loci tested in their study was expressed during early seed development in *A. thaliana*, suggesting that a general silencing of the paternal genome may occur in this tissue. Even if in *A. thaliana* as in maize, no differential methylation correlating with a differential expression of the parental alleles has been reported. There is increasing evidence that DNA methylation is important for the control of imprinting in this model species. The maternal allele of *MEA* is expressed in the female gametophyte and in endosperm associated with torpedo and later stage embryos, while paternally inherited *MEA* alleles are transcriptionally silent in both the young embryo and endosperm. Seeds derived from a mutated maternal allele (*mea*) abort. The partial rescue of the *mea* seed abortion phenotype by fertilisation using pollen from a *ddml* mutant that shows a reduction in overall genomic DNA methylation, indicated that imprinting is lost in these plants (Vielle-Calzada et al., 1999). Loss of imprinting could be a result of changes in chromatin structure or DNA methylation as both are likely to be altered in *ddml* plants.

DNA METHYLATION, CHROMATIN STRUCTURE AND GENE TRANSCRIPTION

The transcriptional machinery in eukaryotic cells functions in the chromatin environment with the template DNA packaged by histone proteins into nucleosomes, 30-nm chromatin fibers, and higher-order structures (Heslop-Harrison, 2000). Recently, the importance of chromosomes and chromatin in the regulation of gene expression has increased with the recognition that chromatin remodelling and histone

acetylation do influence transcription (Martienssen and Henikoff, 1999). A large body of evidence accumulated over the years suggest that DNA methylation has been implicated in the modification of gene expression, and that it is often incompatible with transcriptional activity (Ng and Bird, 1999; Siegfried et al., 1999). However, the inverse correlation between DNA methylation and transcriptional activity has been demonstrated conclusively only for methylation in the promoter regions but not in transcribed parts of genes (Jones, 1999). Is there a connection in the eukaryotic cell between the chromatin structure and DNA methylation for the regulation of gene expression? Experiments performed with in vitro methylated gene sequences revealed that DNA methylation results in the formation of inactive chromatin, and that the silencing effect exerted by methylated cytosines is observed only after the methylated DNA acquired its appropriate chromatin structure (reviewed in Kass et al., 1997; Razin, 1998). Specific transcriptional repressors exist in the form of methyl-CpG-binding proteins (MBPs) that, together with other proteins, originate a repression complex which induces changes in histone acetylation and turns off transcription. Thus, silencing of a gene by methylation involves the generation of a chromatin structure that limits promoter accessibility. In this case, the function of DNA methylation and DNA MTases could be that of maintaining eukaryotic chromosome stability (Smith, 1998). However, the molecular link between the methyl groups on the DNA and the modification of chromatin must be further investigated and clarified.

Finnegan and colleagues (1998) cited the case of the *Pl-Blotched* (*Pl-Bh*) allele of the maize purple plant (*Pl*) gene, as an example in which changes in methylation are secondary to alterations in the chromatin structure. Thus, the primary event appears to be the packaging of chromatin around *Pl-Bh* followed by hypermethylation of *Pl-Bh*. Other evidences that chromatin structure in plants can influence methylation of DNA come from mutation of *MOM* and *DDM1* genes encoding proteins which belong to the SW12/SNF2 family involved in the chromatin-remodelling machinery in cells. Mutations in the human ATRX protein cause some repeats (rDNA) to become hypomethylated, as observed in *ddm1* mutants. As DDM1,

ATRX encodes a SWI/SNF protein, and it was suggested that both plants and animals use chromatin-remodelling engines to control gene silencing (Gibbons et al., 2000). In this context, chromatin remodelling could be a general requirement for replication of condensed, methylated, and inactive regions of the plant genome (Martienssen and Henikoff, 1999).

In all probability DNA methylation, rather than being the primary cause of inactivation, or a secondary consequence of transcriptional inactivation, interacts with chromatin, encouraging promoters that are destined for repression to become even more stably silenced than they would be by the association of chromatin with unmethylated DNA (Bird and Wolffe, 1999).

CONCLUSIONS

DNA methylation is frequently involved in many regulative processes in cells and organisms, and is mainly responsible for the silencing of endogenous, exogenous and/or parasitic DNA. The addition of a methyl group to the cytosine ring is not the solely candidate for the repression of gene expression, because gene silencing events can result from DNA methylation, mutation, a repressive chromatin structure, RNA degradation, or a combination of factors (Selker, 1999). Despite the increasing amount of literature on DNA methylation and its role on gene expression, many questions are still open: i) Why are certain sequences, rather than others, target for DNA methylation? ii) what is the connection between DNA methylation and proteins deputed to chromatin modelling? Is methylation an indirect consequence of a repressive chromatin state that may reinforces gene silencing?

Whatever the answers may be, scientists have to face this phenomenon not only to broaden the basic knowledge on plant genome organisation and regulation, but also for breeding purposes. It is an actual fact that while methylation is frequently associated with gene silencing events in transgenic crops, it also plays a key role in regulating fundamental processes such as embryo development, flowering and seed production.

A last consideration arises from the observation that sequence redundancy, gene inactivation and DNA methylation are mechanisms tightly linked. Since sequence duplication has been proposed as the platform for producing new genes, and considering that methylated cytosines frequently mutate to a thymine (Vairapandi and Duker, 1994), it is conceivable that DNA methylation is an actor of the never ending story of evolution where its role could be the induction of pseudogenes. Pseudogenes, which are non functional genes, are subject to rapid nucleotide changes in the evolutionary process and can become functional again acquiring a new function (Li, 1983). In conclusion it seems not inappropriate also in the case of DNA methylation to cite the statement of Theodosius Dobzhansky: "Nothing in biology makes sense except in the light of evolution."

REFERENCES

- Amedeo, P., Habu, Y., Afsar, K., Mittelsten Scheid, O., and Paszkowski J. (2000) Disruption of the plant gene *MOM* releases transcriptional silencing of methylated genes, *Nature* **405**, 203-206.
- Aniello, F., Locascio, A., Fucci, L., Geraci, G., and Branno, M. (1996) Isolation of cDNA clones encoding DNA methyltransferase of sea urchin *P. lividus*: expression during embryo development, *Gene* **178**, 57-61.
- Antequera, F., and Bird, A.P. (1999) CpG islands as genomic footprints of promoters that are associated with replication origins, *Curr. Biol.* **9**, 661-667.
- Arnaud, P., Goubely, C., Pélissier, T., and Deragon J.M. (2000) The SINE S1 retrotransposon from Brassica can be used in vivo as nucleation centers for de novo methylation, *Abstracts of 6th International Congress of Plant Molecular Biology*, Québec, Canada, June 18-24, 2000, S23-2.
- Assaad, F.F., Tucker, K.L., and Signer, E.R. (1993) Epigenetic repeat induced gene silencing (RIGS) in *Arabidopsis*, *Plant Mol. Biol.* **22**, 1067-1085.
- Banks, J.A., Masson, P., and Fedoroff, N. (1988) Molecular mechanism in the developmental regulation of the maize *Suppressor-mutator* transposable element, *Gen. Dev.* **2**, 1364-1380.
- Barlow, D.P. (1995) Gametic imprinting in mammals, *Science* **270**, 1610-1613.
- Baulcombe, D.C. (1996) Mechanisms of pathogen-derived resistance to viruses in transgenic plants, *Plant Cell* **8**, 1833-1844.
- Baulcombe, D.C. (1999) Fast forward genetics based on virus-induced gene silencing, *Curr. Opin. Plant Biol.* **2**, 109-113.

- Baulcombe, D.C., and English, J.J. (1996) Ectopic pairing of homologous DNA and post-transcriptional gene silencing in transgenic plants, *Curr. Opin. Biotech.* **7**, 173-180.
- Beard, C., Li, E., and Jaenish, R. (1995) Loss of methylation activates *Xist* in somatic but not embryonic cells, *Genes Dev.* **9**, 2325-2334.
- Belanger, F.C., and Hepburn, A.G. (1990) The evolution of CpNpG methylation in plants, *J. Mol. Evol.* **30**, 26-35.
- Bellucci, M., Alpini, A., Paolocci, F., Damiani, F., and Arcioni, S. (1999) Transcription of a maize cDNA in *Lotus corniculatus* is regulated by T-DNA methylation and transgene copy number, *Theor. Appl. Genet.* **98**, 257-264.
- Bender, J., and Fink, G.R. (1995) Epigenetic control of an endogenous gene family is revealed by a novel blue fluorescent mutant of *Arabidopsis*, *Cell* **83**, 725-734.
- Bennetzen, J.L., Schrick, K., Springer, P.S., Brown, W.E., and SanMiguel, P. (1994) Active maize genes are unmodified and flanked by diverse classes of modified, highly repetitive DNA, *Genome* **37**, 565-576.
- Bernacchia, G., Primo, A., Giorgetti, L., Pitto, L., and Cella, R. (1998) Carrot DNA-methyltransferase is encoded by two classes of genes with different pattern of expression, *Plant J.* **13**, 317-329.
- Bestor, T.H. (1992) Activation of mammalian methyltransferase by cleavage of a Zn binding regulatory domain, *EMBO J.* **11**, 2611-2617.
- Bestor, T.H., Laudano, A.P., Mattaliano, R., and Ingram, V.M. (1988) Cloning and sequencing of a cDNA encoding DNA methyltransferase of mouse cells. The carboxyl-terminal domain of the mammalian enzymes is related to bacterial restriction methyltransferases, *J. Mol. Biol.* **203**, 971-983.
- Bestor, T.H., and Verdine, G.L. (1994) DNA methyltransferase, *Curr. Opin. Cell Biol.* **6**, 386-389.
- Bird, A.P. (1995) Gene number, noise reduction and biological complexity, *Trends Genet.* **11**, 94-100.
- Bird, A.P., and Wolffe, A. (1999) Methylation-Induced repression- Belts, Braces, and chromatin, *Cell* **99**, 451-454.
- Brannan, C.I., and Bartolomei, M.S. (1999) Mechanisms of genomic imprinting, *Curr. Opin. Gen. Dev.* **9**, 164-170.
- Brink, R.A. (1960) Paramutation and chromosome organization, *Q. Rev. Biol.* **35**, 120-137.
- Brutnell, T.P., and Dellaporta, S.L. (1994) Somatic inactivation and reactivation of *Ac* associated with changes in cytosine methylation and transposase expression, *Genetics* **138**, 213-225.
- Castilho, A., Neves, N., Rufini-Castiglione, M., Viegas, W., and Heslop-Harrison, J.S. (1999) 5-Methylcytosine distribution and genome organization in Triticale before and after treatment with 5-azacytidine, *J. Cell Sci.* **112**, 4397-4404.
- Cedar, H., Solage, A., Glaser, G., and Razin, A. (1979) Direct detection of methylated cytosine in DNA by use of the restriction enzyme *MspI*, *Nucleic Acids Res.* **6**, 2125-2132.

- Chaudhuri, S., and Messing, J. (1994) Allele-specific parental imprinting of *dzr1*, a posttranscriptional regulator of zein accumulation, *Proc. Natl. Acad. Sci. USA* **91**, 4867-4871.
- Cogoni, C., Ireland, J.T., Schumacher, M., Schmidhauser, T.J., Selker, E.U., and Macino, G. (1996) Transgene silencing of the *al-1* gene in vegetative cells of *Neurospora* is mediated by a cytoplasmic effector and does not depend on DNA:DNA interactions or DNA methylation, *EMBO J.* **15**, 3153-3163.
- Cogoni, C., and Macino, G. (1999a) Homology-dependent gene silencing in plants and fungi: a number of variations on the same theme, *Curr. Opin. Microbiol.* **2**, 657-662.
- Cogoni, C., and Macino, G. (1999b) Gene silencing in *Neurospora crassa* requires a protein homologous to RNA-dependent RNA polymerase, *Nature* **399**, 166-169.
- Colot, V., and Rossignol, J.L., (1999) Eukaryotic DNA methylation as an evolutionary device, *Bioessays* **21**, 402-411.
- Cox, R., and Irving, C.C. (1977) Inhibition of DNA methylation by S-adenosyl methionine with the production of methyl-deficient DNA in regenerating rat liver, *Cancer Res.* **37**, 222-225.
- Crowther, P.J., Cartwright, A.L., Hocking, A., Jefferson, S., Ford, M.D., and Woodcock, D.M. (1989) The effect of *E. coli* host strain on the consensus sequence of regions of the human L1 transposon, *Nucleic acid research* **17**, 7229-7239.
- Dalmay, T., Hamilton, A., Rudd, S., Angell, S., and Baulcombe, D.C. (2000) An RNA-dependent RNA polymerase gene in *Arabidopsis* is required for posttranscriptional gene silencing mediated by a transgene but not by a virus, *Cell* **101**, 543-553.
- Dehio, C., and Schell, J. (1994) Identification of plant genetic loci involved in a posttranscriptional mechanism for meiotically reversible transgene silencing, *Proc. Natl. Acad. Sci. USA* **91**, 5538-5542.
- Diéguez, M.J., Vaucheret, H., Paszkowski, J., and Mittelsten Scheid, O. (1998) Cytosine methylation at CG and CNG sites is not a prerequisite for the initiation of transcriptional silencing in plants but is required for its maintenance, *Mol. Gen. Genet.* **259**, 207-215.
- Elmayan, T., Balzergue, S., Béon, F., Bourdon, V., Daubremet, J., Guènet, Y., Mourrain, P., Palauqui, J.C., Vernhettes, S., Vialle, T., Wostrikoff, K., and Vaucheret, H. (1998) *Arabidopsis* mutants impaired in cosuppression, *Plant Cell* **10**, 1747-1757.
- Elmayan, T., and Vaucheret, H. (1996) Expression of single copies of a strongly expressed 35S transgene can be silenced post-transcriptionally, *Plant J.* **9**, 787-797.
- Elomaa, P., Heliarutta, Y., Griesbach, R.J., Kotilainen, M., Seppanen, P., and Teeri, T.H. (1995) Transgene inactivation in *Petunia hybrida* is influenced by the properties of the foreign gene, *Mol. Gen. Genet.* **248**, 649-656.

- English, J.J., Mueller, E., and Baulcombe, D.C. (1996) Suppression of virus accumulation in transgenic plants exhibiting silencing of nuclear genes, *Plant Cell* **8**, 179-188.
- Fagard, M., and Vaucheret, H. (2000) (Trans)gene silencing in plants: how many mechanisms? *Ann. Rev. Plant Physiol. Plant Mol. Biol.* (in press).
- Federoff, N.V., and Banks, J.A. (1988) Is the *Suppressor-mutator* element controlled by a basic developmental regulatory mechanism? *Genetics* **120**, 559-577.
- Finnegan, E.J. (1989) Eukaryotic transposable elements and genome evolution, *Trends Genet.* **5**, 103-107.
- Finnegan, E.J., and Dennis, E.S. (1993) Isolation and identification by sequence homology of a putative cytosine methyltransferase from *Arabidopsis thaliana*, *Nucleic Acids Res* **21**, 2383-2388.
- Finnegan, E.J., Genger, R.K., Peacock, W.J., and Dennis, E.S. (1998) DNA methylation in plants, *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **49**, 223-247.
- Finnegan, E.J., and Kovac, K.A. (2000) Plant DNA methyltransferase, *Plant Mol. Biol.* **43**, 189-201.
- Finnegan, E.J., Peacock, W.J., and Dennis, E.S. (1996) Reduced DNA methylation in *Arabidopsis thaliana* results in abnormal development, *Proc. Natl. Acad. Sci. USA* **93**, 8449-8454.
- Finnegan, E.J., Peacock, W.J., and Dennis, E.S. (2000) DNA methylation, a key regulator of plant development and other processes, *Curr. Opin. Genet. Dev.* **10**, 217-223.
- Fire, A. (1999) RNA-triggered gene silencing, *Trends Genet.* **15**, 358-363.
- Fire, A., Xu, S., Montgomery, M.K., Kostas, S.A., Driver, S.E., and Mello, C.C. (1998) Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*, *Nature* **391**, 806-811.
- Frediani, M., Giraldi, E., Ruffini Castiglione, M. (1996) Distribution of 5-methylcytosine rich regions in the metaphase chromosomes of *Vicia faba*, *Chromosome Res.* **4**, 141-146.
- Furner, I.J., Sheikh, M.A., and Collett, C.E. (1998) Gene silencing and homology-dependent gene silencing in *Arabidopsis*: genetic modifiers and DNA methylation, *Genetics* **149**, 651-662.
- Frommer, M., McDonald, L.E., Millar, D.S., Collis, C.M., Watt, F., Grigg, G.W., Molloy, P.L., and Paul, C.L. (1992) A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands, *Proc. Natl. Acad. Sci. USA* **89**, 1827-1831.
- Gardiner-Garden, M., Sved, J.A., and Frommer, M. (1992) Methylation sites in angiosperm genes, *J. Mol. Evol.* **34**, 219-230.
- Genger, R.K., Kovac, K.A., Dennis, E.S., Peacock, W.J., and Finnegan, E.J. (1999) Multiple DNA methyltransferase genes in *Arabidopsis thaliana*, *Plant Mol. Biol.* **41**, 269-278.
- Gibbons, R.J., McDowell, T.L., Raman, S., O'Rourke, D.M., Garrick, D., Ayyub, H., and Higgs, D.R. (2000) Mutation in *ATRX*, encoding a SWI/SNF-like protein,

- cause diverse changes in the pattern of DNA methylation, *Nature Genet.* **24**, 368-371.
- Giordano, M., Mattachini, M.E., Cella, R., and Pedrali-Noy, G. (1991) Purification and properties of a novel DNA methyltransferase from cultured rice cells. *Biochem. Biophys. Res. Commun.* **177**, 711-719.
- Goubely, C., Arnaud, P., Tatout, C., Heslop-Harrison, J.S., and Deragon, J.M. (1999) S1 SINE retroposons are methylated at symmetrical and non-symmetrical positions in *Brassica napus*: identification of a preferred target site for asymmetrical methylation, *Plant Mol. Biol.* **39**, 243-255.
- Grant, S.R. (1999) Dissecting the mechanisms of posttranscriptional gene silencing: divide and conquer. *Cell* **96**, 303-306.
- Green, M.M. (1988) Mobile elements and spontaneous gene mutation, in M.E. Lambert, J.F. McDonald, and I.B. Weinstein, (eds), *Eukaryotic transposable elements as mutagenic agents*, Cold Spring Harbor NY, Cold Spring Harbor Laboratory Press, pp.41-50.
- Gruenbaum, Y., Naveh-Man, T., Cedar, H., and Razin, A. (1981) Sequence specificity of methylation in higher plant DNA, *Nature* **292**, 860-862.
- Henikoff, S., and Comai, L. (1998) A DNA methyltransferase homolog with a chromodomain exists in multiple polymorphic forms in *Arabidopsis*, *Genetics* **149**, 307-318.
- Heslop-Harrison, J.S. (2000) Comparative genome organization in plants: from sequence and markers to chromatin and chromosomes, *Plant Cell* **12**, 617-635.
- Hirochika, H. (1997) Retrotransposons of rice: their regulation and use for genome analysis, *Plant Mol. Biol.* **35**, 231-240.
- Hirochika, H., and Hirochika, R. (1993) *Ty1-copia* group retrotransposons as ubiquitous components of plant genomes, *Jpn. J. Genet.* **68**, 35-46.
- Hirochika, H., Okamoto, H., and Kakutani, T. (2000) Silencing of retrotransposons in *Arabidopsis* and reactivation by the *ddm1* mutation, *Plant Cell* **12**, 357-369.
- Holliday, R., and Pugh, J.E., (1975) DNA modification mechanisms and gene activity during development, *Science* **187**, 226-232.
- Hsieh, C.L., and Lieber, M.R. (1992) CpG methylated minichromosomes become inaccessible for V(D)J recombination after undergoing replication, *EMBO J* **11**, 315-325.
- Hsu, D.W., Lin, M.J., Lee, T.L., Wen, S.C., Chen, X., and James Shen, C.K. (1999) Two major forms of DNA (cytosine-5) methyltransferase in human somatic tissues, *Proc. Natl. Acad. Sci. USA* **96**, 9751-9756.
- Iglesias, V.A., Moscone, E.A., Papp, I., Neuhuber, F., Michalowski, S., Phelan, T., Spiker, S., Matzke, M., and Matzke, A.J.M. (1997) *Plant Cell* **9**, 1251-1264.
- Ingelbrecht, I., Van Houdt, H., Van Montagu, M., and Depicker, A. (1994) Posttranscriptional silencing of reporter transgenes in tobacco correlates with DNA methylation, *Proc. Natl. Acad. Sci. USA* **91**, 10502-10506.
- Isles, A.R., and Wilkinson, L.S. (2000) Imprinted genes, cognition and behaviour, *Trends Cognitive Sci.* **4**, 309-318.

- Jacobsen, S.E., and Meyerowitz, E.M. (1997) Hypermethylated SUPERMAN epigenetic alleles in *Arabidopsis*, *Science* **277**, 1100-1103.
- Jacobsen, S.E., Sakai, H., Finnegan, E.J., Cao, X., and Meyerowitz, E.M. (2000) Ectopic hypermethylation of flower-specific genes in *Arabidopsis*, *Curr. Biol.* **10**, 179-186.
- Jeddeloh, J.A., Bender, J., and Richards, E.J. (1998) The DNA methylation locus DDM1 is required for maintenance of gene silencing in *Arabidopsis*, *Gen. Dev.* **12**, 1714-1725.
- Jeddeloh, J.A., and Richards, E.J. (1996) ^mCCG methylation in angiosperms, *Plant J.* **9**, 579-586.
- Jeddeloh, J.A., Stokes, T.L., and Richards, E.J. (1999) Maintenance of genomic methylation requires a SW12/SNF2-like protein, *Nature Genet.* **22**, 94-97.
- Jones, A.L., Thomas, C.L., and Maule, A.J. (1998) De novo methylation and co-suppression induced by a cytoplasmically replicating plant RNA virus, *EMBO J.* **17**, 6385-6393.
- Jones, P.A. (1999) The DNA methylation paradox, *Trends Genet.* **15**, 34-37.
- Jones, P.A., and Laird, P.W. (1999) Cancer epigenetics comes of age, *Nature Genet.* **21**, 163-167.
- Jones, P.A., and Taylor, S.M. (1980) Cellular differentiation, cytidine analogs and DNA methylation, *Cell* **20**, 85-93.
- Jorgensen, R. (1994) Developmental significance of epigenetic imposition in plant genome: a paragenetic function of chromosomes, *Devel. Genetics* **15**, 523-532.
- Jost, J.P., and Saluz, H.P. (1993) *DNA methylation: Molecular Biology and Biological Significance*, Birkhauser, Basel.
- Kakutani, T., Jeddeloh, J., Flowers, S., Munakata, K., and Richards, E.J. (1996) Developmental abnormalities and epimutations associated with DNA hypomethylation mutations, *Proc Natl Acad Sci USA* **93**, 12406-12411.
- Kakutani, T., Munakata, K., Richards, E.J., and Hirochika, H. (1999) Meiotically stable inheritance of DNA hypomethylation induced by *ddm1* mutation of *Arabidopsis thaliana*, *Genetics* **151**, 831-838.
- Kass, S., Pruss, D., and Wolffe, A. (1997) How does DNA methylation repress transcription? *Trends Genet.* **13**, 444-449.
- Kennerdell, J., and Carthew, R. (1998) *Drosophila* *frizzled* and *frizzled 2* act in the wingless pathway as determined by dsRNA-mediated genetic interference, *Cell* **95**, 1017-1026.
- Kermicle, J.L., and Alleman, M. (1990) Gametic imprinting in maize in relation to the angiosperm life cycle *Dev. Suppl.* 9-14.
- Kumaputra, S.P., and Hall T.C. (1998) Longevity of 5-azacytidine-mediated gene expression and re-establishment of silencing in transgenic rice, *Plant Mol. Biol.* **38**, 1113-1122.
- Lei, H., Oh, S.P., Okano, M., Juttermann, R., Goss, K.A., Jaenisch, R., and Li, E. (1996) De novo DNA cytosine methyltransferase activities in mouse embryonic stem cells, *Development* **122**, 3195-3205.

- Li, E., Berad, C., and Jaenisch, R. (1993) Role for DNA methylation in genomic imprinting, *Nature* **366**, 362-365.
- Li, E., Bestor, T.H., and Jaenisch, R. (1992) Targeted mutation of the DNA methyltransferase gene results in embryonic lethality, *Cell* **69**, 915-926.
- Li, W.H. (1983) Evolution of duplicate genes and pseudogenes, in M. Nei and R.K. Koehn (eds.) *Evolution of genes and proteins*, Sinauer Assoc. Sunderland, pp. 14-37.
- Lindbo, J.A., Silva-Rosales, L., Proebsting, W.M., and Dougherty (1993) Induction of highly specific antiviral state in transgenic plants: implications for regulation of gene expression and virus resistance, *Plant Cell* **5**, 1749-1759.
- Liu, W.M., Maraia, R.J., Rubin, C.M., and Schmid, C.W. (1994) Alu transcripts: cytoplasmic localisation and regulation by DNA methylation, *Nucl. Acids Res.* **22**, 1087-1095.
- Liu, W.M., and Schmid, C.W. (1993) Proposed roles for DNA methylation in *Alu* transcriptional repression and regulation by DNA methylation, *Nucl. Acids Res.* **21**, 1351-1359.
- Luff, B., Pawlowski, L., and Bender, J. (1999) An inverted repeat triggers cytosine methylation of identical sequences in *Arabidopsis*, *Mol Cell* **3**, 505-511.
- Lund, G., Ciceri, P., and Viotti, A. (1995a) Maternal-specific demethylation and expression of specific alleles of zein genes in the endosperm of *Zea mays* L., *Plant J.* **8**, 571-581.
- Lund, G., Messing, J., and Viotti, A. (1995b) Endosperm-specific demethylation and activation of specific alleles of α -tubulin genes of *Zea mays* L., *Mol. Gen. Genet.* **6**, 716-722.
- Luo, M., Bilodeau, P., Koltunow, A., Dennis, E.S., Peacock, W.J., and Chaudhury, A.M. (1999) Genes controlling fertilization-independent seed development in *Arabidopsis thaliana*, *Proc. Natl. Acad. Sci. USA* **96**, 296-301.
- Maloisel, L., and Rossignol, J.L. (1998) Suppression of crossing-over by DNA methylation in *Ascobolus*, *Genes Dev* **12**, 1381-1389.
- Martienssen, R. (1998) Chromosome imprinting in plants. *Curr. Opin. Gen. Dev.* **8**, 240-244.
- Martienssen, R., and Henikoff, S. (1999) The *house & garden* guide to chromatin remodelling, *Nature Gen.* **22**, 6-7.
- Matassi, G., Melis, R., Kuo, K.C., Macaya, G., Gehrke, C.W., and Bernardi, G. (1992) Large-scale methylation patterns in the nuclear genomes of plants, *Gene* **122**, 239-245.
- Matzke, A.J.M. and Matzke, M.A. (1998) Position effects and epigenetic silencing of plant transgenes, *Curr. Opin. Plant Biol.* **1**, 142-148.
- Matzke, A.J.M., Neuhuber, F., Park, Y.D., Ambros, P.F., and Matzke, M.A. (1994) Homology-dependent gene silencing in transgenic plants: Epistatic silencing loci contain multiple copies of methylated transgenes, *Mol. Gen. Genet.* **244**, 219-229.
- Matzke, M.A., Matzke, A.J.M. (1995) How and why do plants inactivate homologous (trans)genes? *Plant Physiol.* **107**, 679-685.

- Matzke, M.A., and Matzke, A.J.M. (1998) Epigenetic silencing of plant transgenes as a consequence of diverse cellular defense responses, *Cell Mol. Life Sci.* **5**, 94-103.
- Matzke, M.A., Matzke, A.J.M., and Eggleston, W.B. (1996) Paramutation and transgene silencing: a common response to invasive DNA? *Trends Plant Sci.* **1**, 382-388.
- McClelland, M. (1983) The frequency and distribution of methylatable DNA sequences in leguminous plant protein coding genes, *J. Mol. Evol.* **19**, 346-354.
- Melquist, S., Luff, B., and Bender, J. (1999) *Arabidopsis* PAI gene arrangements, cytosine methylation and expression, *Genetics* **153**, 401-413.
- Mette, M.F., van der Winden, J., Matzke, M.A., and Matzke, A.J.M. (1999) Production of aberrant promoter transcripts contributes to methylation and silencing of unlinked homologous promoters *in trans*, *EMBO J.* **18**, 241-248.
- Metzlaff, M., O'Dell, M., Cluster, P.D., and Flavell, R.B. (1997) RNA-mediated RNA degradation and chalcone synthase A silencing in *Petunia*, *Cell* **88**, 845-854.
- Meyer, P., and Heidmann, I. (1994) Epigenetic variants of a transgenic petunia line show hypermethylation in transgene DNA: an indication for specific recognition of foreign DNA in transgenic plants, *Mol. Gen. Genet.* **243**, 390-399.
- Meyer, P., Niedenhof I., and Ten Lohuis, M. (1994) Evidence for cytosine methylation of non-symmetrical sequences in transgenic *Petunia hybrida*, *EMBO J.* **13**, 2084-2088.
- Meyer, P., Saedler, H. (1996) Homology-dependent gene silencing in plants, *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **47**, 23-48.
- Mittelsten Scheid, O., Afsar, K., and Paszkowski, J. (1998) Release of epigenetic gene silencing by trans-acting mutation in *Arabidopsis*, *Proc. Natl. Acad. Sci. USA* **95**, 632-637.
- Mohandas, T., Sparkes, R.S., and Shapiro, L.J. (1981) Reactivation of an inactive human X chromosome: evidence for X inactivation by DNA methylation, *Science* **211**, 393-396.
- Montgomery, M.K., and Fire, A. (1998) Double-stranded RNA as a mediator in sequence-specific genetic silencing and co-suppression, *Trends Genet.* **14**, 255-258.
- Mourrain, P., Béclin, C., Elmayan, T., Feuerbach, F., Godon, C., Morel, J.B., Jouette, D., Lacombe, A.M., Nikic, S., Picault, N., Rémoûé, K., Sanial, M., Vo, T.A., and Vaucheret, H. (2000) *Arabidopsis* SGS2 and SGS3 genes are required for posttranscriptional gene silencing and natural virus resistance, *Cell* **101**, 533-542.
- Nan, X., Campoy, J., and Bird, A. (1997) MeCP2 is a transcriptional repressor with abundant binding sites in genomic chromatin, *Cell* **88**, 471-481.
- Napoli, C., Lemieux, C., and Jorgensen, R.A. (1990) Introduction of a chimeric chalcone synthase gene into petunia results in reversible co-suppression of homologous gene *in trans*, *Plant Cell* **2**, 279-289.
- Ng, H.H., and Bird, A. (1999) DNA methylation and chromatin modification, *Curr. Opin. Gen. Dev.* **9**, 158-163.

- Nick, H., Bowen, B., Ferl, R.J., and Gilbert, W. (1986) Detection of cytosine methylation in the maize alcohol dehydrogenase gene by genomic sequencing, *Nature* **319**, 243-246.
- Noyer-Weidner, M., and Trautner, T.A. (1993) Methylation of DNA in prokaryotes, in J.P. Jost, and H.P. Saluz (eds) *DNA methylation: Molecular Biology and Biological Significance*, Birkhauser, Basel, pp.39-108.
- Oakeley, E.J., and Jost, J.P. (1996) Non-symmetrical cytosine methylation in tobacco pollen DNA, *Plant Mol. Biol.* **31**, 927-930.
- Oakeley, E.J., Podestà, A., and Jost, J. (1997) Developmental changes in DNA methylation of the two tobacco pollen nuclei during maturation, *Proc. Natl. Acad. Sci. USA* **94**, 11721-11725.
- Okano, M., Bell, D.W., Haber, D.A., and Li, E. (1999) DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development, *Cell* **99**, 247-257.
- Okano, M., Xie, S., and Li, E. (1998) Cloning and characterization of a family of novel mammalian DNA (cytosine-5) methyltransferases, *Nat. Genet.* **19**: 219-220.
- Okazaky, Y., Okuizumi, H., Takada, S., Takahara, T., and Hayashizaki, Y. (1997) Protocols for RLGS gel production (Chapter 3), in Y. Hayashizaki and S. Watanabe (eds.), *Restriction Landmark Genome Scanning (RLGS)*, Springer-Verlag, Tokyo, pp 17-36.
- Park, Y.D., Papp, I., Moscone, E.A., Iglesias, V.A., Vaucheret, H., Matzke, A.J.M., and Matzke, M.A. (1996) Gene silencing mediated by promoter homology occurs at the level of transcription and results in meiotically heritable alterations in methylation and gene activity, *Plant J.* **9**, 183-194.
- Paro, R., and Harte, P.J. (1996) The role of polycomb group and thrithorax group chromatin complexes in the maintenance of determined cell states, in V.E.A. Russo, R.A. Martienssen, and A.D. Riggs (eds), *Epigenetic Mechanism of gene regulation*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, pp. 507-528.
- Pelissier, T., Thalmeir, S., Kempe, D., Sanger, H.L., and Wassenegger, M. (1999) Heavy de novo methylation at symmetrical and non-symmetrical sites is a hallmark of RNA-directed DNA methylation, *Nucleic Acids Res.* **27**, 1625-1634.
- Pradhan, S., and Adams, R.L.P. (1995) Distinct CG and CNG DNA methyltransferases in *Pisum sativum*, *Plant J.* **7**, 471-481.
- Pradhan, S., Bacolla, A., Wells, R.D., and Roberts, R.J. (1999) Recombinant human DNA (cytosine-5) methyltransferase. I. Expression, purification, and comparison of de novo and maintenance methylation. *J. Biol. Chem.* **274**, 33002-33010.
- Pradhan, S., Cummings, M., Roberts, R.J., and Adams, R.L.P. (1998) Isolation, characterization and baculovirus-mediated expression of the cDNA encoding cytosine DNA methyltransferase from *Pisum sativum*, *Nucleic Acids Res.* **26**, 1214-1222.
- Prols, F., and Meyer, P. (1992) The methylation patterns of chromosomal integration regions influence gene activity of transferred DNA in *Petunia hybrida*, *Plant J.* **2**, 465-475.

- Que, Q., Wang, H.Y., English, J.J., and Jorgensen, R.A. (1997) The frequency and degree of co-suppression by sense chalcone synthase transgenes are dependent on transgene promoter strength and are reduced by premature nonsense codons in the transgene coding sequence, *Plant Cell* **9**, 1357-1368.
- Ratcliff, F.G., MacFarlane, S.A., and Baulcombe, D.C. (1999) Gene silencing without DNA: RNA-mediated cross-protection between viruses, *Plant Cell* **11**, 1207-1215.
- Raugei, G., Bazzicalupo, M., Federici, M., Gallori, E., Pepino, R., and Polsinelli, M. (1981) Effect of a new pyrimidine analog on *B.subtilis* growth, *J. Bacteriology* **145**, 1079-1083.
- Razin, A. (1998) CpG methylation, chromatin structure and gene silencing-a three way connection, *EMBO J.* **17**, 4905-4908.
- Richards, E.J. (1997) DNA methylation and plant development, *Trends Genet.* **13**, 319-323.
- Riggs, A.D. (1975) X-inactivation, differentiation and DNA methylation, *Cytogenet. Cell Genet.* **14**, 9-25.
- Ronchi, A., Petroni, K., and Tonelli, C. (1995) The reduced expression of endogenous duplications (REED) in the maize R gene family is mediated by DNA methylation, *EMBO J.* **14**, 5318-5328.
- Ronemus, M.J., Galbiati, M., Ticknor, C., Chen, J.C., and Dellaporta, S.L. (1996) Demethylation-induced developmental pleiotropy in *Arabidopsis*, *Science* **273**, 654-657.
- Rossignol, J.L., and Faugeron, G. (1994) Gene inactivation triggered by recognition between DNA repeats, *Experientia* **50**, 307-317.
- Ruiz, F., Vayssie, L., Klotz, C., Sperling, L., and Madeddu, L. (1998) Homology-dependent gene silencing in *Paramecium*, *Mol. Biol. Cell* **9**, 931-943.
- Russinova, E., and de Vries, S. (2000) Parental contribution of plant embryos, *Plant Cell* **12**, 461-463.
- Schiebel, W., Pélissier, T., Riedel, L., Thalmeir, S., Schiebel, R., Kempe, D., Lottspeich, F., Sanger, H.L., and Wassenegger, M. (1998) Isolation of an RNA-directed RNA polymerase-specific cDNA clone from tomato, *Plant Cell* **10**, 2087-2101.
- Schlappi, M., Raina, R., and Fedoroff, N. (1994) Epigenetic regulation of the maize Spm transposable element: novel activation of a methylated promoter by TnpA, *Cell* **77**: 427-437.
- Selker, E.U. (1997) Epigenetic phenomena in filamentous fungi: useful paradigms or repeat-induced confusion? *Trends Genet.* **13**, 296-301.
- Selker, E.U. (1999) Gene silencing: repeats that count, *Cell* **97**, 157-160.
- Selker, E.U., Fritz, D.Y., and Singer, M.J. (1993) Dense nonsymmetrical DNA methylation resulting from repeat-induced point mutation in *Neurospora*, *Science* **262**, 1724-1728.
- Shapiro, H.S., (1970) in H.S. Sauer (ed), *CRC Handbook of Biochemistry Selected Data for Molecular Biology*, The Chemical Rubber Co., Ohio, pp.80-103.

- Shapiro, H.S. (1976) Distribution of purines and pyrimidines in deoxyribonucleic acids, in G.D. Fasman (ed), *CRC Handbook of Biochemistry and Molecular Biology* vol. 2, 3rd edn. CPC Press, Cleveland, pp241-281.
- Sheldon, C.C., Rouse D.T., Finnegan, E.J., Peacock, W.J., and Dennis, E.S. (2000) The molecular basis of vernalization: the central role of FLOWERING LOCUS C (FLC), *Proc. Natl. Acad. Sci. USA* **97**, 3753-3758.
- Sigfried, Z., Eden, S., Mendelsohn, M., Feng, X., Tsuberi, B.Z., and Cedar, H. (1999) DNA methylation represses transcription in vivo, *Nature Gen.* **22**, 203-206.
- Smardon, A., Spoerke, J.M., Stacey, S.C., Klein, M.E., Mackin, N., and Maine E.M. (2000) EGO-1 is related to RNA-directed RNA polymerase and functions in germ-line development and RNA interference in *C. elegans*, *Curr. Biol.* **10**, 169-178.
- Smith, S.S. (1998) Stalling of DNA methyltransferase in chromosome stability and chromosome remodelling, *Int. J. Mol. Med.* **1**, 147-156.
- Stam, M., Mol, J.N.M., and Kooter, J.M. (1998) Position-dependent methylation and transcriptional silencing of transgenes in inverted T-DNA repeats: implications for posttranscriptional silencing of homologous host genes in plants, *Mol. Cell Biol.* **18**, 6165-6177.
- Steimer, A., Amedeo, P., Afsar, K., Fransz, P., Mittelsten Scheid, O., and Paszkowski J (2000) Endogenous targets of transcriptional gene silencing in *Arabidopsis*, *Plant Cell* **12**, 1165-1178.
- Stein, R., Gruenbaum, Y., Pollack, Y., Razin, A., and Cedar, H. (1982) Clonal inheritance of the pattern of DNA methylation in mouse cells, *Proc. Natl. Acad. Sci. USA* **79**, 61-65.
- Stokes, T.L., and Richards, E.J. (2000) Mum's the word: MOM and modifiers of transcriptional gene silencing, *Plant Cell* **12**, 1003-1006.
- Sudarsanam, P., and Winston, F. (2000) The Swi/Snf family nucleosome-remodelling complexes and transcriptional control, *Trends Genet.* **16**, 345-551.
- Ten Louis, M., Muller, A., Heidmann, I., Niedenhof, I., and Meyer, P. (1995) A repetitive DNA fragment carrying a hot spot for de novo methylation enhances expression variegation in tobacco and petunia, *Plant J.* **8**, 919-932.
- Theiss, G., Schleicher, R., Schimpff-Weiland, R., and Follmann, H. (1987) Dna methylation in wheat, *Eur J Biochem* **167**, 89-96.
- Thierry, D., and Vaucheret, H. (1996) Sequences homology requirements for transcriptional silencing of 35S transgenes and post-transcriptional silencing of nitrite reductase (trans)genes by the tobacco 271 locus, *Plant Mol Biol.* **32**, 1075-1083.
- Vairapandi, M., and Duker, N.J. (1994) Excision of ultraviolet-induced photoproducts of 5-methylcytosine from DNA, *Mutat. Res.* **315**, 85-94.
- Van der Krol, A.R., Mur, L.A., Beld, M., Mol, J.N.M., and Stuitji, A.R. (1990) Flavonoid genes in petunia: Addition of a limited number of gene copies may lead to a suppression of gene expression, *Plant Cell* **2**, 291-299.

- Vanyushin, B.F., Belozersky, A.N., Kokurina, N.A., and Kadirova, D.X. (1968) 5-methylcytosine and 6-methylaminopurine in bacterial DNA, *Nature* **218**, 1066-1067.
- Vanyushin, B.F., Tkacheva, S.G., and Belozersky, A.N. (1970) Rare base in animal DNA, *Nature* **225**, 948-949.
- Vaucheret, H. (1993) Identification of a general silencer for 19S and 35S promoters in a transgenic tobacco plant: 90 bp of homology in the promoter sequences are sufficient for trans-inactivation, *CR Acad. Sci. Paris* **316**, 1471-1483.
- Vielle-Calzada, J.P., Baskar, R., and Grossniklaus, U. (2000) Delayed activation of the parental genome during seed development, *Nature* **404**, 91-94.
- Vielle-Calzada, J.P., Thomas, J., Spillane, C., Coluccio, A., Hoepfner, M.A., and Grossniklaus, U. (1999) Maintenance of genomic imprinting at the *Arabidopsis medea* locus requires zygotic *DDM1* activity, *Genes Dev.* **13**, 2971-2982.
- Vongs, A., Kakutani, T., Martienssen, A., and Richards, E.J. (1993) *Arabidopsis thaliana* DNA methylation mutants, *Science* **260**, 1926-1928.
- Wagner, I., and Capesius, I. (1981) Determination of 5-methylcytosine from plant DNA by high-performance liquid chromatography, *Biochim. Biophys. Acta* **654**, 52-56.
- Wang, L.H., Heinlein, M., and Kunze, R. (1996) Methylation pattern of Activator transposase binding sites in maize endosperm, *Plant Cell* **8**, 747-758.
- Wang, R.Y., Gehrke, C.W., and Ehrlich, M. (1980) Comparison of bisulfite modification of 5-methyldeoxycytidine and deoxycytidine residues, *Nucleic Acids Res.* **8**, 4777-4790.
- Wassenegger, M., Heimes, S., Riedel, L., and Sanger, H.L. (1994) RNA-directed de novo methylation of genomic sequences in plants, *Cell* **76**, 567-576.
- Wassenegger, M., and Pélissier, T. (1998) A model for RNA-mediated gene silencing in higher plants, *Plant Mol. Biol.* **37**, 349-362.
- Waterhouse, P.M., Graham, M.W., and Wang, M.B. (1998) Virus resistance and gene silencing in plants can be induced by simultaneous expression of sense and antisense RNA, *Proc. Natl. Acad. Sci. USA* **95**, 13959-13964.
- Wassenegger, M. (2000) RNA-directed DNA methylation, *Plant Mol. Biol.* **43**, 203-220.
- Waterhouse, P.M., Smith, N.A., and Wang, M.B. (1999) Virus resistance and gene silencing: killing the messenger, *Trends Plant Sci.* **4**, 452-457.
- Wendel, J.F. (2000) Genome evolution in polyploids, *Plant Mol. Biol.* **42**, 225-249.
- Wessler, S.R., Bureau, T.E., and White, S.E. (1995) LTR-retrotransposon and MITEs: important players in the evolution of plant genomes, *Curr. Opin. Genet. Dev.* **5**, 814-821.
- Wigler, M., Levy, D., and Peruchio, M. (1981) The somatic replication of DNA methylation, *Cell* **24**, 33-40.
- Wolfe, A.P., Jones, P.L., and Wade, P.A. (1999) DNA demethylation, *Proc. Natl. Acad. Sci. USA* **96**, 5894-5896.

- Xie, S., Wang, Z., Okano, M., Nogami, M., Li, Y., He, W.W., Okumura, K., and Li, E. (1999) Cloning, expression and chromosome locations of the human DNMT3 gene family, *Gene* **236**, 87-95.
- Ye, F., and Signer, E.R. (1996) RIGS (repeat-induced gene silencing) in *Arabidopsis* is transcriptional and alters chromatin configuration, *Proc. Natl. Sci. Acad. USA* **93**, 10881-10886.
- Yoder, J.A., Walsh, C.P., and Bestor, T.H. (1997) Cytosine methylation and the ecology of intragenomic parasites, *Trends Genet.* **13**, 335-340.
- Yoder, J.A., Yen, R.W., Vertino, P.M., Bestor, T.H., and Baylin, S.B. (1996) New 5' regions of the murine and human genes for DNA (cytosine-5)-methyltransferase, *J. Biol. Chem.* **271**, 31092-31097.

QTL MAPPING IN CROP PLANTS

SVEN BODE ANDERSEN and ANNA MARIA TORP

*The Royal Veterinary and Agricultural University,
Dept. Agricultural Sciences, Thorvaldsensvej 40,
1871 Frederiksberg C. Denmark.*

INTRODUCTION

Since rediscovery of Mendelian genetics by the beginning of the 20th century, studies of the effect of mutations in genes have provided a profound contribution to our understanding of biology. Results of such studies have founded the general idea of phenotype as the result of interactions between genes and environment. Refined methodology for studies of single genes with major effect on phenotype range from detailed segregation studies to cloning, sequencing, expression studies, and genetic transformation. It has enabled detailed understanding of the genetic component of many major traits in plant breeding. This knowledge has had strong influence on modern agriculture through introduction of traits like reduced plant height (dwarfism), disease resistance, and high quality traits, and much more is to be expected with recent molecular developments. Understanding of the genetic component is a strong platform for studies of environmental effects and dissection of the underlying molecular events. Therefore, qualitative genetics in many instances is the key to our understanding of plant biology. This paper deals with a comparable development in tools to study genetics of quantitative traits.

TRANSITION FROM QUALITATIVE TO QUANTITATIVE TRAITS

One of the very first experiences following reinvention of Mendelian genetics was the observation that many traits apparently could not be

analysed with the new laws of genetics. Most traits of practical importance like plant height, yield, and many types of disease resistance did not form discrete classes in their offspring and therefore genes in the Mendelian way could not be used for their description. It was still clear, however, that genes were involved because pure breeding lines with high yield or high resistance consistently produced offspring with high yield or resistance, while the opposite was the case for low yielding or susceptible material.

Nilsson-Ehle (1909) and Johannsen (1909) were among the first to construct a scientific explanation to this phenomenon, while still maintaining Mendelian genetics. Major amendments to the “one gene one trait” theory were that several different genes can affect the same trait and that environment affects such traits in addition to the genes. Several genes affecting one trait have the well known effect that offspring segregation becomes complex with many classes, which may not be distinguishable and whose segregation ratios can be difficult to figure out. This phenomenon is illustrated in Figure 1 top row for the case of one, two, and three genes affecting a disease resistance. Effects of growth environment on segregation patterns of such traits are that genotype classes become blurred or indistinguishable in the offspring if environmental effects are strong (Figure 1). This is because genetically identical plants are phenotypically different due to variable environmental growing conditions.

Most traits in plants important for agriculture or other purposes actually show a type of offspring segregation where phenotypic classes cannot be distinguished. Such traits are generally termed quantitative, to distinguish them from qualitative traits, which give a clear segregation into distinct offspring classes. Therefore genetic constitution of a plant cannot be derived from its phenotype. Individual genes for such traits cannot be distinguished using traditional offspring analysis

QUANTITATIVE GENETICS OR BIOMETRICS

Because it was not possible to describe the behaviour of quantitative trait segregation with simple Mendelian models, alternative theories

named Quantitative genetics have been developed during the 20th century. Quantitative genetics uses a statistical approach to describe simultaneously the effects of all genes and environmental factors as well as their interactions. With suitable experimental approaches and statistical modelling the total variation in the offspring can be divided into components ascribed to 1) effects of genes, 2) effects of environmental factors and 3) their interactions (G x E). Genetic components may be further divided into components due to additive and non-additive effects, which can sometimes be divided further into components for dominance effects and interactions between gene loci (epistasis). Also interactions between different lines and different environments can be analysed in considerable detail. For detailed description of quantitative genetics the reader is referred to excellent text books like: Hill et al. (1998), Falconer (1989), Mather and Jinks (1982).

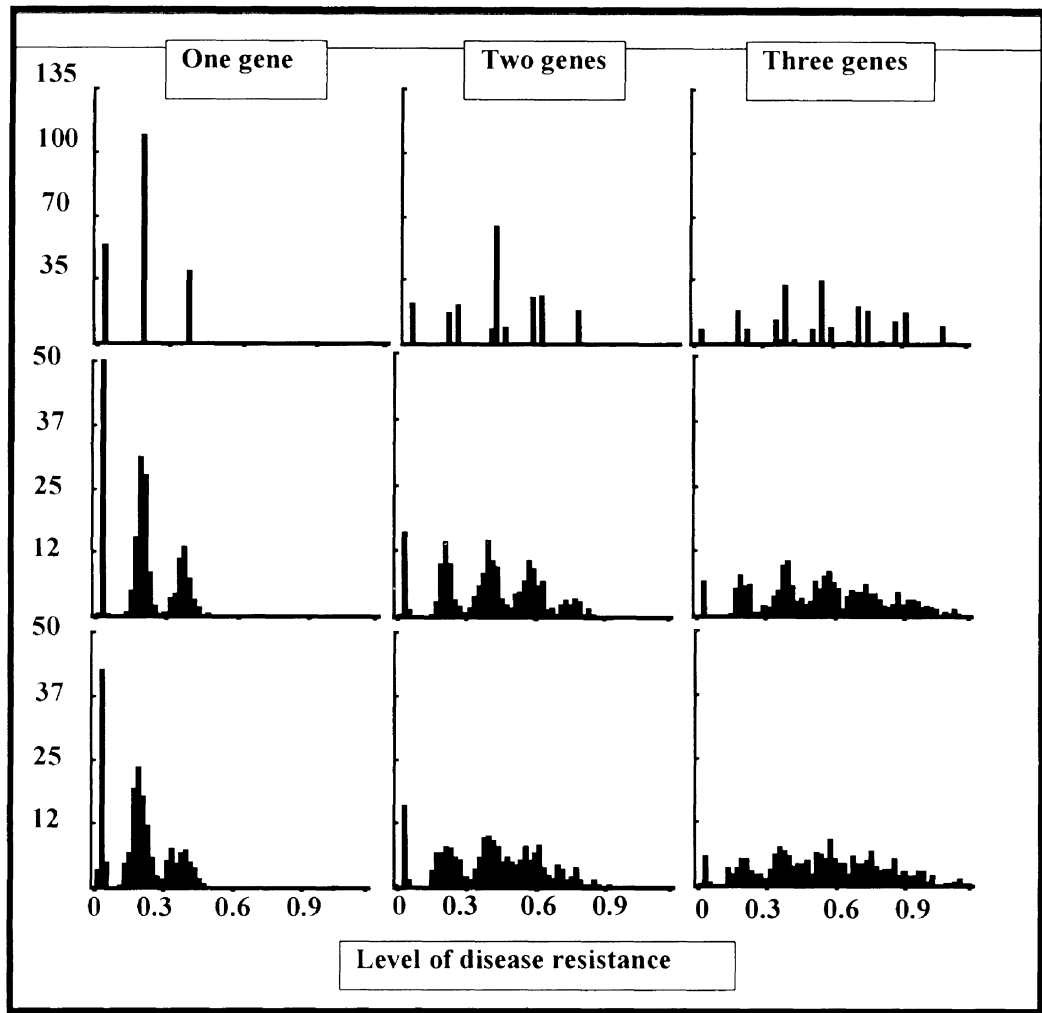


Figure 1. Segregation in F_2 progeny of a disease resistance affected by one, two or three genes (columns) and by increasing environmental effects (rows: 2, 10, 25 percent environmental effect, respectively). Ordinate: number of plants at each level of resistance, abscissa: relative score of disease resistance. Data simulated with the program Supergene (Andersen 2000).

Since effects of individual genes cannot be distinguished it was necessary to assume in quantitative genetics that all genes affecting the trait have the same effect, *i.e.*, an average gene. Furthermore, to take advantage of the power and generality of normal theory it has been a general assumption that a large number of such average genes are involved. Methods have been developed to detect and model the effects of linkage, but they are in general not very powerful. Therefore most models also assume that all genes involved in the trait are inherited independently without linkage.

In spite of these necessary assumptions, which obviously will not always be fulfilled, quantitative genetic theory and methods have had a major impact on plant breeding during the 20th century. Among the achievements should be mentioned the use of estimates of heritability (broad and narrow sense) for most traits of major importance to choose between different strategies in line and population breeding, theories for understanding of heterosis/inbreeding depression, use of combining ability in hybrid breeding, and methods for analysis of GxE interactions for selection of genotypes with stable performance in different environments.

With new tools in genetic science providing large number of genetic markers potentially covering entire plant genomes, it is gradually becoming possible to dissect quantitative traits into their underlying genes named quantitative trait loci (QTLs). With QTL mapping it is possible to locate to chromosome areas the major genes affecting a quantitative trait, and to study their interaction with other genes affecting the trait, as well as their interaction with environmental factors. With these tools an improved description of inheritance of quantitative traits is possible.

APPROACHES TO QTL MAPPING

All approaches to chromosomal mapping of genes affecting quantitative traits use some sort of genetic marker, which theoretically can be of any type: Morphological, protein/isozymes or DNA based markers. However, in most cases a large number of such genetic markers are needed to cover the chromosomes. With the development of systems based on analysis of small DNA segments such markers have become available in very large numbers. This has been a major reason for rapid development in both theories and practical approaches to QTL mapping during recent years.

Another prerequisite for QTL mapping is some type of offspring segregating for the quantitative trait. Mapping populations are generally produced through initial hybridisation of parents with different genes for the trait to be studied. In the case of heterozygous parental material the F₁ offspring will segregate and may be used directly as mapping population. In other cases further generation offspring is needed among which the simplest populations to produce

are F_2 or BC_1 . Except in cases where plants can be evaluated for the trait after cloning, a major disadvantage of such heterozygous mapping populations is that each genotype is available as only a single plant. This often does not permit exact measurement of quantitative traits. The obstacle may to some extent be compensated for by the large number of individuals that can be rapidly obtained with these types of populations. However, very large numbers of individuals to be analysed put strain on the marker analysis part of the projects. Instead, self-pollinated progenies from each original plant in heterozygous mapping populations may be evaluated for the trait in replicated trials to provide better estimates of phenotypic effects. Dominance as well as some types of epistatic effects, however, can only be studied with heterozygous populations. The best choice of mapping population will be either chromosome doubled haploids (DH) or recombined inbred (RI) lines. Although they take more time or efforts to produce they consist of homozygous true breeding lines, which can be easily multiplied to allow evaluation for the trait in many environments. Dominant or co-dominant markers are equally informative with such homozygous material. Genetic segregation in these two types of mapping populations is almost identical. RI populations will show slightly higher recombination compared to DH populations because of recombination in several cycles of meioses during RI formation compared with recombination in only one meiosis in the case of DH lines. Use of DH mapping populations are by far the simplest and they will be used as the material for illustration of principles throughout this chapter. Theories and methods for the remaining types of mapping populations are more complex, but the basic principles are unchanged.

Single marker associations

The simplest approaches to mapping of QTLs on chromosomes use single marker associations. Association of single markers with measurements of a quantitative trait has been known for a long time (Sax 1923) and has been used as arguments for the existence of single genes affecting quantitative traits.

Figure 2 illustrates single marker association based on simulated data assuming one gene (QH/QL) affecting disease resistance linked to a genetic marker (MH/ML). This disease resistance is affected by other genes in addition to the Q-gene and by environmental factors.

Therefore, the DH offspring from an F_1 hybrid between a highly resistant and a susceptible homozygous parent, as illustrated in Figure 2a, will result in a continuous distribution with respect to disease resistance (Figure 2b). Linkage between the marker MH/ML and the resistance gene QH/QL means that the marker allele MH from the high resistant parent will in most cases follow the QH allele from the resistant parent. Accordingly, the marker allele ML from the low resistant parent will in most cases be inherited together with the QL allele of the resistance gene from the low resistant parent. Segregation of the resistance gene cannot be observed because genotypic classes are indistinguishable but segregation of the linked marker can be observed. Therefore, DH offspring from the hybrid can be sorted into two groups according to the MH (Figure 2c) or ML (Figure 2d) marker alleles

If the marker M is located inside the resistance gene then the sorting according to M alleles will also provide complete sorting for resistance alleles. In most cases, however, there will be a distance between the genetic marker and the QTL, which is normally expressed by means of the recombination frequency r . Among the DH offspring carrying marker alleles MH the majority with a frequency of $(1-r)$ will also carry the allele for high resistance (QH). However, a minority with frequency r among these lines have experienced genetic recombination between the marker and the QTL and therefore inherit the low resistance allele (QL), as illustrated in Figure 2c. Therefore, each of the two marker classes of offspring plants consist of a mixture of two underlying distributions with respect to the quantitative trait, one distribution due to parental and one distribution due to recombined gametes.

It is possible to model the mixed distributions e.g. assuming that they are all normal distributions. The resistance measurement Y for each plant in the offspring can be assigned a likelihood given its marker allele. For plants carrying marker allele MH this would be:

$$L(Y | MH): (1-r)\Psi(Y, \mu_h, \sigma_h) + r \Psi(Y, \mu_L, \sigma_L)$$

where r is the recombination frequency between marker and QTL, and $\Psi(Y, \mu, \sigma)$ is the normal Likelihood function with expectation μ and standard deviation σ calculated for measurement Y . The two standard

errors may contain both variation due to other genes affecting the trait as well as environmental effects.

The first term models parental and the second recombinant DH genotypes. Similarly, likelihood of observations for lines carrying the low resistance marker allele can be modelled:

$$L(Y | ML): r \Psi(Y, \mu_h, \sigma_h) + (1-r) \Psi(Y, \mu_L, \sigma_L)$$

A total likelihood for all plants in the offspring can be constructed as the product sum of likelihood for each plant and optimised with respect to the five parameters in the model: $r, \mu_h, \mu_L, \sigma_h, \sigma_L$ to get their maximum likelihood estimates. Maximum likelihood estimates from single marker associations generally have large standard errors, however, and the chance of detecting the QTL decreases rapidly with increasing recombination distance (Weller 1986, Luo and Kearsey 1989).

Single marker association methods are generally used simply to check whether unlinked markers show an association with the quantitative trait in question. If the marker in question is linked to one of the genes affecting the quantitative trait then sorting the offspring according to marker groups should also lead to some sorting of the material according to the trait. Therefore, with association between a marker and a quantitative trait, means of the quantitative measurement of the two marker groups become different. In the simulated example of Figure 2, a simple t-test can be performed to see whether there is significant difference in average disease resistance between the two marker groups. A more general approach to single marker association test will use a linear regression of trait measurements on marker genotypes assuming no dominance. It is not possible, however, with such an approach to estimate correctly either the recombination distance between marker and QTL or the effect of the QTL, *i.e.*, allele substitution value ($\mu_h - \mu_L$). A QTL with large effect may show small difference between the two marker groups if recombination between marker and the QTL is high, while a QTL with small effect may show a relatively large difference between marker groups if the marker is tightly linked. Also with this approach, the chance of detecting the QTL is reduced quickly with increasing distance between marker and QTL (Lander and Botstein 1989).

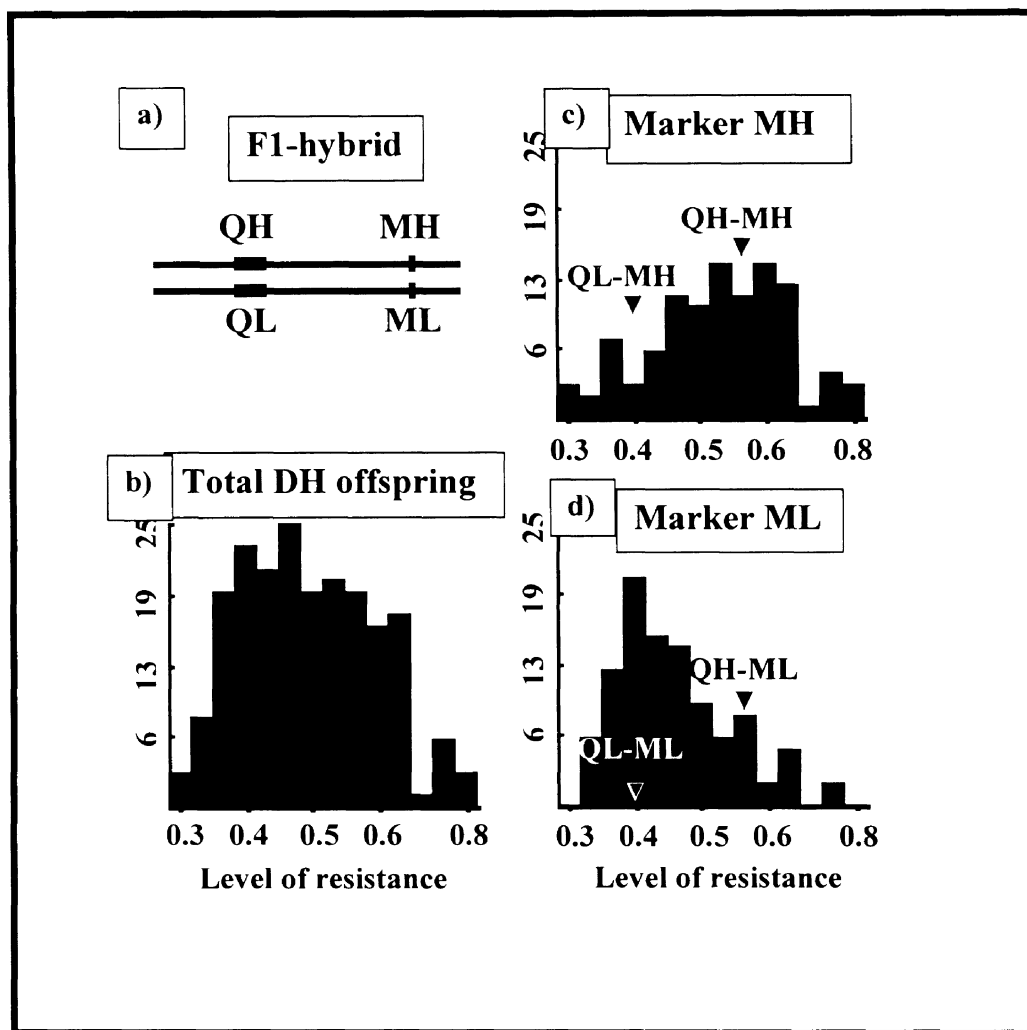


Figure 2. Simulated segregation of a quantitative disease resistance with one gene QH/QL linked to a genetic marker MH/ML. a) Linkage situation in F₁ hybrid. Distribution of disease resistance in total DH offspring (b) and the two marker allele classes (c and d). Simulated with the program Supergene (Andersen 2000).

A general problem of multiple testing is immediately faced if single marker tests are used in large numbers to see whether one or more genetic markers in a large collection are associated with a quantitative trait. If many markers are tested with a trait using the standard 5 percent significance level, then on average five percent of such tests will indicate association, even if there is no association between the trait and any of the markers at all (Type I error). In principle the level of significance in multiple tests can be increased to assure for instance that overall probability of one false positive is 5 percent. However,

increase of the significance level also decreases sensitivity of the test. Hence fewer true associations are detected (Type II error). There is no simple mathematical way for calculation of the best significance level, as these tests are normally not independent.

A general empirical approach based on re-sampling, to the calculation of optimal significance levels for testing with genetic markers have been described by Churchill and Doerge (1994). If marker genotypes for each individual line are retained in the data, while measurements of the quantitative trait for each genotype are randomly permuted, then the true associations between markers and trait will be broken. In reality a situation conforming with the null hypothesis of no associations is created. From thousand or more such repeated calculations with permuted trait values the actual distribution of the test variable under the null hypothesis can be determined and the level of the test variable corresponding to *e.g.* 5 percent overall false positives can be determined for the particular set of data.

Interval mapping

A much better approach for mapping of QTLs than the single marker associations discussed above will use information from two linked markers preferably on each side of the QTL (*i.e.*, flanking markers) as illustrated on top of Figure 3. With segregation of two markers among DH offspring it is possible to divide the offspring into four different marker classes: Lines that inherited both marker alleles from the low resistant parent (Figure 3a), lines with both markers from the high resistant parent (Figure 3b), and two groups (Figure 3c and d) in which each line received one marker allele from the high resistant and one marker allele from the low resistant parent.

As can be seen from the simulated data in Figure 3, distribution of the quantitative trait becomes quite different in the four different marker classes. Each of these marker classes consists of the mixture of two different distributions, one for lines that carry the high resistant allele (QH) and one for those lines that inherited the low resistance allele (QL). Lines that received both marker alleles from the same original parent (Figure 3 a and b) are mostly parental types with the low resistance allele (Figure 3a) or the high resistance allele (Figure 3b). Both these groups however, can also contain a low frequency of lines with the alternate resistance allele, due to rare double crossover on

both sides of the resistance gene. In Figure 3a these double recombinants are

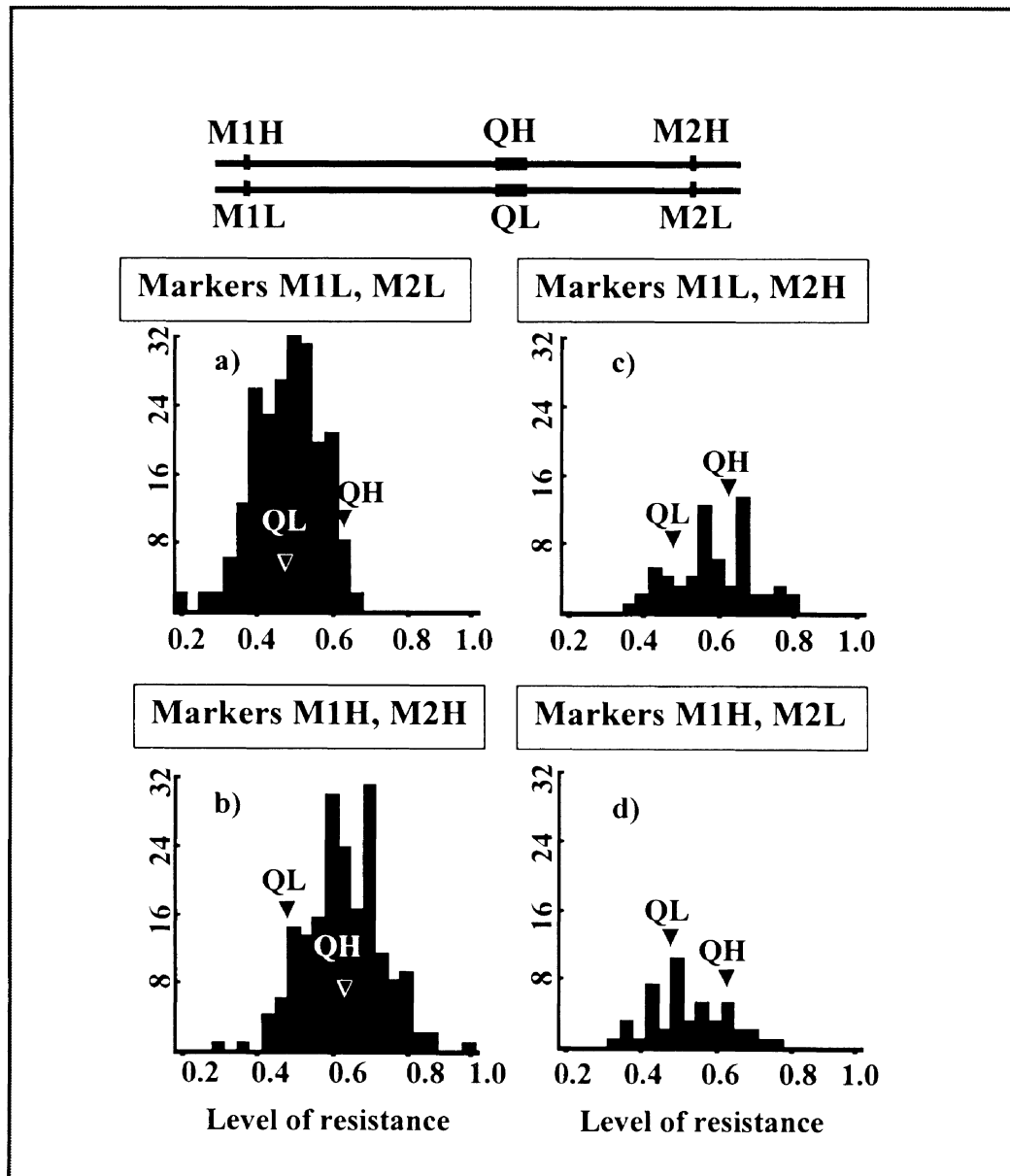


Figure 3. Simulated segregation of a quantitative disease resistance with one gene QH/QL linked to two flanking markers M1 and M2, both with alleles from the resistant (M1H, M2H) and the susceptible parent (M1L, M2L). Linkage situation in F_1 on top. Distribution of disease resistance in marker allele classes of the DH offspring (a-d). Simulated with the program Supergene (Andersen 2000).

marked QH while they are marked QL in Figure 3b. Provided the double recombinants are rare the average difference in resistance between these two parental marker classes is an approximate estimate

of the average difference in resistance between lines with one or the other resistance allele (allele substitution value).

The two mixed distributions in each of the recombined marker classes (Figure 3c and d) are the result of genetic recombination taking place either in the first chromosome segment (M1-Q) or the second segment (Q-M2). In Figure 3c plants resulting from recombination in the first segment (M1-Q) received the high resistance allele (QH) and form a distribution with higher average resistance than those plants that obtained QL after recombination in Q-M2. In the same way, in Figure 3d, resistance values for recombinants of the first chromosome segment (M1-Q) are concentrated around the expected resistance value (QL) of the low resistance allele, while resistance values for recombinants in the second chromosome segment (Q-M2) concentrate around the expected resistance level (QH) of the high resistance allele. Relative length of the two chromosome segments (M1-Q and Q-M2) is thus reflected in the relative number of plants in the two mixed distributions in these recombined marker classes.

Like in the case of single marker associations it is possible to model the mixed distributions in each of the four marker classes with appropriate normal distributions. Jensen (1989) described a method to estimate all parameters in a model without prior assumptions on marker QTL sequences or recombination distances. All three recombination distances can be estimated as well as expected means for the two quantitative trait locus alleles, two variances for the fitted normal distributions and a parameter to allow for distorted segregation of the QTL alleles. Much more widely used methods, however, make the assumption that the QTL is situated between the two markers (flanking markers) and the recombination frequency between these flanking markers is assumed known, *i.e.*, the genetic map for localisation of a QTL is assumed fixed. If the recombination frequency between the flanking markers is assumed known and denoted r then recombination frequency r_1 in the first segment (M1-Q) and r_2 in the second segment (Q-M2, Figure 3) relate to r as: $r=r_1+r_2-2r_1r_2$, *i.e.*, no interference is assumed. For each resistance value measured in the four marker classes, a likelihood conditional on parameters and marker classes can be calculated as:

$$L(Y | M1L, M2L): [[(1-r_1)(1-r_2)]N_{low} + r_1r_2N_{high}]/(1-r)$$

$$L(Y \mid M1H, M2H): [r_1 r_2 N_{low} + [(1-r_1)(1-r_2)] N_{high}] / (1-r)$$

$$L(Y \mid M1L, M2H): [(1-r_1)r_2 N_{low} + r_1(1-r_2) N_{high}] / r$$

$$L(Y \mid M1H, M2L): [r_1(1-r_2) N_{low} + (1-r_1)r_2 N_{high}] / r$$

N_{high} and N_{low} are the normal likelihood functions for each of the two resistance allele classes: $N_{high} = \Psi(Y, \mu_h, \sigma_h)$ and $N_{low} = \Psi(Y, \mu_l, \sigma_l)$, Y is the measured resistance value of the genotype, μ_h and μ_l are expected resistance values for plants carrying the high or low resistance allele and σ_h and σ_l are the standard deviations of the two distributions. Divisors r and $(1-r)$ are needed to make frequencies of genotypes within each marker class sum to unity for conditioning on flanking markers. The most interesting parameters of this flanking marker association problem to be estimated are the effect of the QTL, $(\mu_h - \mu_l)$ and the position on the chromosome of the QTL given by r_1 and r_2 .

A widely used method designated “interval mapping” was introduced by Lander and Botstein (1989) and described in more detail by Van Ooijen (1992). The above problem was simplified assuming that variation is identical in the different distributions ($\sigma_h = \sigma_l$), which is often a good approximation particularly with small QTL effects. The QTL is assumed positioned at small distances (e.g. one cM) between the flanking markers. For each position the corresponding r_1 and r_2 can be calculated and from them the mixing proportion between the two distributions in the mixture of each marker class. Subsequently, the two normal distributions are fitted using non-linear regression to obtain μ_h , μ_l and σ for each putative QTL location together with the likelihood of the fitted model. To assess the fit of the model at each location the likelihood of a similar model without any QTL is fitted. A LOD-score for each point is calculated as the \log_{10} of the ratio of likelihood from model with QTL to the likelihood of model without QTL. By plotting the LOD-score for each interval against the genome map a QTL likelihood map is constructed for the entire chromosome (Lander and Botstein 1989; Van Ooijen 1992). When maxima in the profile exceed a predetermined threshold, a QTL is declared to be present at that location. Threshold values of LOD-scores for acceptance of a QTL are often set to 3, which means that the

likelihood from model with a QTL is 1000 times higher than likelihood from the model without a QTL. A more general approach is to empirically estimate the threshold for a particular set of data using re-sampling (Churchill and Doerge 1994). Lander and Botsteins interval mapping approach has been implemented in the software package Mapmaker/QTL (Paterson et al. 1988).

Another method to estimate the parameters of the flanking marker association with a quantitative trait uses regression of trait measurements on their marker class expectations (Haley and Knott, 1992, Martinez and Curnow, 1992). Expectations of resistance measurements in the marker class M1L, M2L, can be modelled as:

$$E(M1L, M2L): \mu_0 + [(1-r_1)(1-r_2)](\mu_0 - \mu_l) + r_1 r_2 (\mu_h - \mu_0) / (1-r)$$

where μ_0 is the expected mean of all resistance measurements.

With $(\mu_h - \mu_0) = -(\mu_l - \mu_0) = \beta$ the equation simplifies into a simple linear regression of the trait measurements on their relative expectations:

$$E(M1L, M2L): \mu_0 + \beta[r_1 r_2 - (1-r_1)(1-r_2)] / (1-r)$$

$$E(M1H, M2H): \mu_0 + \beta[(1-r_1)(1-r_2) - r_1 r_2] / (1-r)$$

$$E(M1L, M2H): \mu_0 + \beta[r_1(1-r_2) - (1-r_1)r_2] / r$$

$$E(M1H, M2L): \mu_0 + \beta[(1-r_1)r_2 - r_1(1-r_2)] / r$$

Fitting the model uses an interval mapping approach as above. The QTL is assumed positioned at regular intervals along the chromosome. For each location with known r_1 and r_2 the coefficients for β can be calculated as relative expectations. The two parameters μ_0 and β can subsequently be estimated with a simple linear regression of trait measurements on the relative expectations. Likelihood and LOD-scores can be obtained from the residual sum of squares and plotted against chromosome positions to locate areas with high LOD-scores. Alternatively the error sum of squares from fitted models is plotted in which case a minimum error will indicate location of a QTL. Regression of trait values on their relative expectations is computationally much faster than the interval mapping methods used by Lander and Botstein (1989) and they produce almost identical

results in most situations. Whittaker et al. (1996) presented a further simplification of the above regression approach which permits localisation of the QTL based on regression of trait measurements directly on marker genotypes of the two flanking markers.

The interval mapping methods do not provide natural standard errors for estimates of position (*i.e.*, r_1 and r_2) of the QTL on the chromosome segment. Lander and Botstein (1989) suggested the use of LOD support intervals; *i.e.*, intervals delimited by the positions on either side of the estimated QTL location corresponding to a decrease in LOD-score of 1 or 2 LOD-units. Other methods to obtain confidence intervals are based on simulations (Mangin et al. 1994; Mangin and Goffinet 1997) or bootstrapping (Visscher et al. 1996).

Marker Regression

Association between multiple markers linked to one QTL is a further development of the ideas behind single marker association to include information from all surrounding markers in the linkage group. by Kearsey and Hyne (1994) and Hyne and Kearsey (1995) described principles for analysis of QTLs using information from all markers in a linkage group.

Imagine a situation of several markers in a linkage group surrounding one QTL as illustrated for the hybrid in Figure 4a. If the position of the QTL is assumed known then each marker in the linkage group has a well defined recombination distance r from the QTL affecting the quantitative trait. For each genetic marker down the chromosome we can then sort the offspring according to marker classes, *i.e.*, those lines that have the marker allele from the parent with high resistance and those lines that have the allele from the low resistance parent in the particular genetic marker. For the class that has the marker allele from the high resistance parent, we can write the expectation of its means X_h and similar for the alternative marker class:

$$\text{Mean } (X_h) = (1 - r)\mu_h + r \mu_l$$

$$\text{Mean } (X_l) = r \mu_h + (1 - r)\mu_l$$

The first term for X_h denotes lines without recombination between marker and QTL so the line has inherited the high resistance allele of

the QTL with trait value μ_h . The second term comprises recombinants with a frequency of r , which received the low resistance allele with value μ_l . Expectation of resistance in the group of lines that received the marker allele from the low resistant parent is derived using similar arguments.

For each of the markers down the chromosome then the difference (D) between mean measurement of the quantitative trait in the two marker groups can be modelled as:

$$D(X_h - X_l) = (1 - 2r) \mu_h - (1 - 2r) \mu_l = (1 - 2r) (\mu_h - \mu_l)$$

This means that we can regress the differences of quantitative trait measurements for each marker $D(X_h - X_l)$ locus against $(1 - 2r)$ calculated for that particular marker. An estimate of the allelic substitution value $(\mu_h - \mu_l)$ of the QTL can be obtained as the slope of the regression through origin (Figure 4b) and the fit of the model is measured as the error sum of squares.

To find the most likely position of a quantitative trait locus on the chromosome segment the QTL is assumed positioned at regular small intervals along the chromosome and the regression model fitted for each position. The most likely position of the QTL is where the best fit of the model (minimum error sum of squares) with the data is obtained. Since with this approach only one model is fitted to each linkage group, the existence of a QTL in the area can be tested using conventional F or t statistics. If N different linkage groups are searched for QTLs, they can often be assumed to provide independent tests, which means that the significance level for the test statistics can be corrected to provide e.g. an overall risk of 5% for one false positive in the entire analysis. From the linear regression in addition a standard error is obtained for the estimate of the effect of the QTL (allelic substitution value).

Multiple simultaneous QTL mapping

All the above approaches using interval mapping, regression on marker class means or marker regression may have problems if more than one major QTL is segregating in the mapping population. If two QTLs are linked on the same chromosome segment they may both

affect measurements in the same direction and may be detected together as only one QTL. Alternatively, they may counteract each other so none of them are detected (Haley and Knott 1992; Martinez and Curnow 1992; Jansen 1994). QTLs segregating in other parts of the genome may distort distributions of the marker classes, particularly, if small mapping populations are used. This can lead to biased

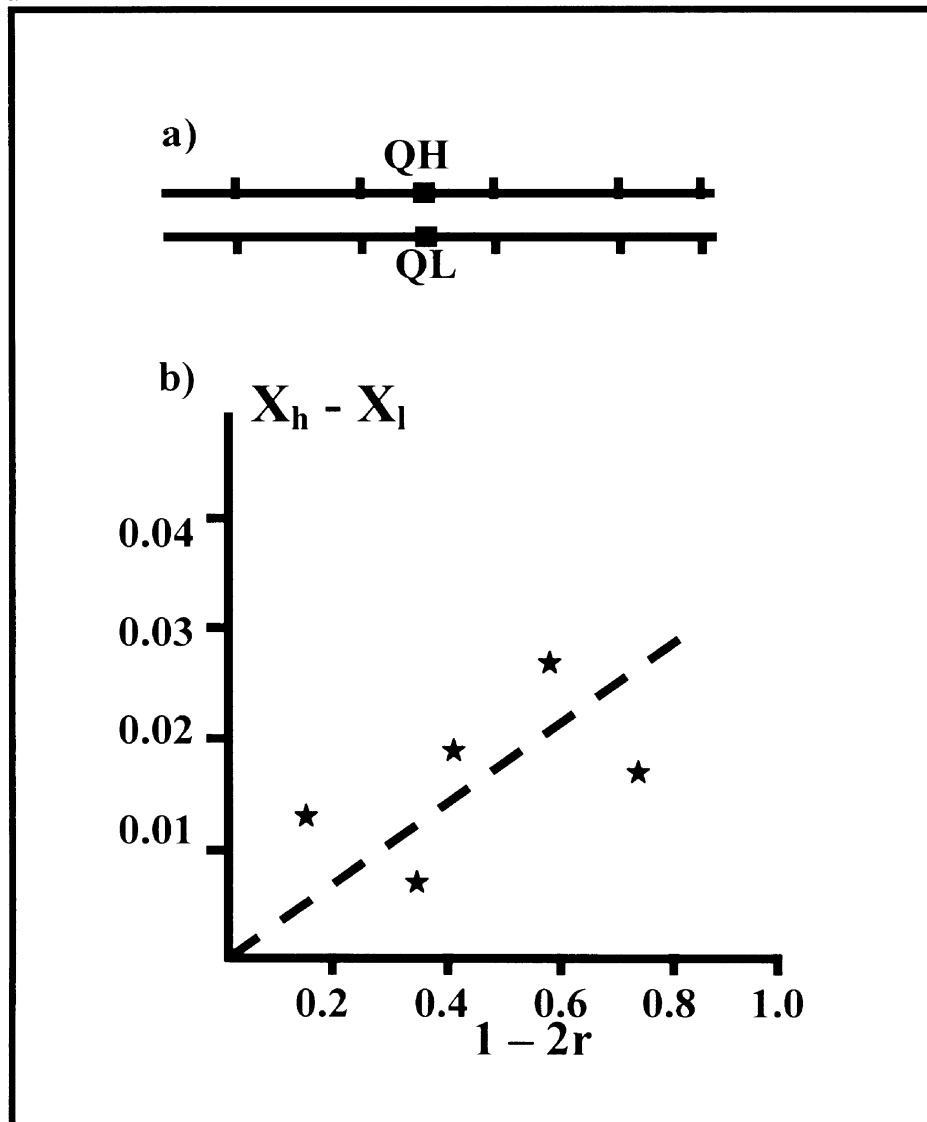


Figure 4. Marker regression.

estimates, QTLs passing undetected or even detection of non-existing QTLs. Such situations with two or several QTLs segregating at the same time are to be expected when mapping genes underlying

quantitative traits. Two major types of approaches are used to overcome these problems.

A widely used approach, Multiple QTL Mapping (MQM) (Jansen 1993, Jansen and Stam 1994) or Composite Interval Mapping (Zeng 1993, 1994) consists of a two step procedure. The first step use some approach, e.g., stepwise regression, to identify genetic markers in the data, which can explain a significant part of the variation for the trait, when used as simple co-factors in a multiple regression model. The linkage map of markers is subsequently searched one linkage group at a time for the existence of a QTL with cofactors fitted simultaneously. During the search of a particular marker interval, cofactors flanking the interval are removed from the regression model while other cofactors are maintained. Inclusion of marker cofactors in the model results in an increased power of detection of QTLs, because the cofactors absorb part of the residual variance from other segregating QTLs in the genome. Several computer programs are available to perform composite interval mapping including QTL-cartographer (Basten et al. 1997), MapQTL (van Ooijen and Maliapaard 1996), PlabQTL (Utz and Melchinger 1996), and MQTL (Tinker and Mather 1995). Efficiency of some of these strategies have been studied using computer simulation and a simple four step strategy for statistical testing has been proposed (Visscher et al. 2000)

Another major approach takes a stepwise strategy to the construction of models with multiple simultaneously fitted QTLs (Haley and Knott 1992; Hyne and Kearsey 1995). All linkage groups are initially searched for QTLs. The QTL with the highest LOD-score found is then included into the model and the remaining variation not explained by the model is searched for QTLs in all linkage groups. The QTL with the highest LOD-score from each cycle of search is included into the model until no more significant improvement in likelihood is obtained. Multiple regression of trait values on marker scores for identification of pairs of markers with significant regression coefficients of similar sign (Whittaker et al. 1996) is especially suitable for this type of multiple QTL mapping.

Finally, Monte Carlo methods optimising likelihood (Jansen 1996) or using Bayesian approach (Satagopan et al. 1996) have been proposed for simultaneously estimating multiple QTLs. Such approaches are

still computationally heavy, but they may be further improved with the development of more powerful computers in the future.

IMPROVEMENTS TO QUANTITATIVE TRAIT DESCRIPTION USING QTL MAPPING

The rapid development of theory and methods for detection of quantitative trait loci during the last few decades has inspired a great deal of enthusiasm within genetics and plant science. It is now possible to identify and chromosome locate major genes underlying quantitative traits and to study their individual properties and interactions. Recently, there has been a tremendous increase in the number of QTLs identified for important traits including: Earliness and photoperiod response (Maheswaran et al. 2000, Sourdille et al. 2000), complex inherited types of disease resistance (Waldron et al. 1999, Nilsson et al. 1999, Qi et al. 1999, van Heusden et al. 1999), quality traits like malting quality (Marquez-Cedillo et al. 2000), tolerance to or escape from drought (Sari-Gorla et al. 1999, Crasta et al. 1999, Courtois et al. 2000) or the ability to stay green during periods of water stress (Xu et al. 2000, Tao et al. 2000).

It is clear, however, that statistical methods for detection of multiple simultaneously segregating QTLs still need improvements. Improved techniques for likelihood optimisation of existing types of models using genetic algorithms (Carlborg et al. 2000) may allow much more complex models to be fitted in the future. Further development of Bayesian approaches for QTL mapping may improve the handling of non-normal data (Yi and Xu 2000) and dominant markers in heterozygous populations (Satagopan, et al. 1996). Improvements in detection and parameter estimation during QTL analysis will lead to improved descriptions of main effects of QTLs for important traits as well as modelling of gene x environment and gene x gene (epistasis) interactions (Crossa et al. 1999, Wang et al. 1999). Such a development combined with further technical improvements may lead to efficient marker assisted breeding programs in the future (Young 1999, Hospital et al. 2000).

CONCLUSION AND PROSPECTS

Increase in computational capacity during the last few decades has triggered rapid improvements to theory and practice of QTL mapping. Major improvements comprise interval mapping using flanking markers analysed with maximum likelihood or regression approaches as well as marker regression using information from all markers in a linkage group. These techniques have greatly improved both detection of QTLs as well as estimation of their effects and chromosomal position. Also for simultaneous detection of multiple QTLs, major progress has been achieved through the use of marker co-factors and multiple QTL modelling. The development has enabled detailed dissection of quantitative traits into underlying QTLs in many plant species and further rapid development can be expected in the near future.

REFERENCES

- Andersen S.B. (2000): Virtual plant breeding with the program Supergene. www.supergene.dk
- Basten, C.J., Weir, B.S., Zeng, Z.-B. (1997): QTL Cartographer: A Reference Manual and Tutorial for QTL Mapping. Department of Statistics, North Carolina State University, Raleigh, NC.
- Carlborg, O., Andersson, L., Kinghorn, B. (2000): The use of a genetic algorithm for simultaneous mapping of multiple interacting quantitative trait loci. *Genetics* 155: 2003-2010.
- Churchill, G.A., Doerge, R.W. (1994): Empirical threshold values for quantitative trait mapping. *Genetics* 138: 963-971.
- Courtois, B., McLaren, G., Sinha, P.K., Prasad, K., Yadav, R., Shen, L. (2000): Mapping QTLs associated with drought avoidance in upland rice. *Mol. Breeding* 6: 55-66.
- Crasta, O.R., Xu, W.W., Rosenow, D.T., Mullet, J., Nguyen, H.T. (1999): Mapping of post-flowering drought resistance traits in grain sorghum: association between QTLs influencing premature senescence and maturity. *Mol. Gen. Genet.* 262: 579-588.
- Crossa, J., Vargas, M., van Eeuwijk, F.A., Jiang, C., Edmeades, G.O., Hoisington, D. (1999): Interpreting genotype x environment interaction in tropical maize using linked molecular markers and environmental covariables. *Theor. Appl. Genet.* 99: 611-625.
- Falconer, D.S. (1989): *Introduction to Quantitative Genetics*, 3rd edn. Longman, Harlow.
- Haley, C.S., Knott, S.A. (1992): A simple regression method for mapping quantitative trait loci in line crosses using flanking markers. *Heredity* 69: 315-324.
- Hill, J., Becker, H.C., Tigerstedt, P.M.A. (1998): *Quantitative and ecological aspects of plant breeding*. Chapman & Hall, London.
- Hospital, F., Goldringer, I., Openshaw, S. (2000): Efficient marker-based recurrent selection for multiple quantitative trait loci. *Genet. Res.* 75: 357-368.
- Hyne, V., Kearsey, M.J. (1995): QTL analysis: further uses of 'marker regression'. *Theor. Appl. Genet.* 91: 471-476.

- Jansen, R.C. (1993): Interval mapping of multiple quantitative trait loci. *Genetics* 135: 205-211.
- Jansen, R.C. (1994): Controlling the type I and type II errors in mapping quantitative trait loci. *Genetics* 138: 871-881.
- Jansen, R.C. (1996): A general Monte Carlo method for mapping multiple quantitative trait loci. *Genetics* 142: 305-311.
- Jansen, R.C., Stam, P. (1994): High resolution of quantitative traits into multiple loci via interval mapping. *Genetics* 136: 1447-1455.
- Jensen, J. (1989): Estimation of recombination parameters between a quantitative trait locus (QTL) and two marker gene loci. *Theor. Appl. Genet.* 78: 613-618.
- Johannsen, W. (1909): *Elemente der exakten Erblchkeitslehre*. Fisher, Jena.
- Kearsey, M.J., Hyne, V. (1994): QTL analysis: a simple 'marker-regression' approach. *Theor. Appl. Genet.* 89: 698-702.
- Lander, E.S., Botstein, D. (1989): Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics* 121: 185-199.
- Luo, Z.W., Kearsey, M.J. (1989): Maximum likelihood estimation of linkage between a marker gene and a quantitative locus. *Heredity* 63: 401-408.
- Maheswaran, M., Huang, N., Sreerangasamy, S.R., McCouch, S.R. (2000): Mapping quantitative trait loci associated with days to flowering and photoperiod sensitivity in rice (*Oryza sativa* L.). *Mol. Breeding*, 6: 145-155.
- Mangin, B., Goffinet, B., Rebaï, A. (1994): Constructing confidence intervals for QTL location. *Genetics* 138: 1301-1308.
- Mangin, B., Goffinet, B. (1997): Comparison of several confidence intervals for QTL location. *Heredity* 78: 345-353.
- Marquez-Cedillo, L.A., Hayes, P.M., Jones, B.L., Kleinhofs, A., Legge, W.G., Rosnagel, B.G., Sato, K., Ullrich, E., Wesenberg, D.M. (2000): QTL analysis of malting quality in barley based on the doubled-haploid progeny of two elite North American varieties representing different germplasm groups. *Theor. Appl. Genet.* 101:173-184.
- Martinez, O., Curnow, R.N. (1992): Estimating the locations and the sizes of the effects of quantitative trait loci using flanking markers. *Theor. Appl. Genet.* 85: 480-488.
- Mather, K., Jinks, J.L. (1982): *Biometrical Genetics*, 3rd edn. Chapman & Hall, London.
- Nilsson-Ehle, H. (1909): *Kreuzungsuntersuchungen an Hafer und Weizen*. Lund.
- Nilsson, N.O., Hansen, M., Panagopoulos, A.H., Tuveusson, S., Ehlde, M., Christiansson, M., Rading, I.M., Rissler, M., Kraft, T., (1999): QTL analysis of *Cercospora* leaf spot resistance in sugar beet. *Plant Breeding* 118: 327-334.
- Paterson, A., Lander, E., Lincoln, S., Hewitt, J., Peterson, S., Tanksley, S. (1988): Resolution of quantitative traits into mendelian factors using a complete RFLP linkage map. *Nature* 335: 721-726.
- Qi, X., Jiang, G., Chen, W., Nix, R.E., Stam, P., Lindhout, P., (1999): Isolate-specific QTLs for partial resistance to *Puccinia hordei* in barley. *Theor. Appl. Genet.* 99: 877-884.
- Sari-Gorla, M., Krajewski, P., Di Fonzo, N., Villa, M., Frova, C. (1999): Genetic analysis of drought tolerance in maize by molecular markers. II. Plant height and flowering. *Theor. Appl. Genet.* 99: 289-295.
- Satagopan, J.M., Yandell, B.S., Newton, M.A., Osborn, T.C. (1996): A Bayesian approach to detect quantitative trait loci using markov chain Monte Carlo. *Genetics* 144: 805-816.
- Sax, K., (1923): The association of size differences with seed-coat pattern and pigmentation in *Phaseolus vulgaris*. *Genetics* 8: 552-560.
- Sourdille, P., Snape, J.W., Cadalen, T., Charmet, G., Nakata, N., Bernard, S., Bernard, M. (2000): Detection of QTLs for heading time and photoperiod response in wheat using a doubled-haploid population. *Genome* 43: 487-494.
- Tao, Y.Z., Henzell, R.G., Jordan, D.R., Butler, D.G., Kelly, A.M., McIntyre, C.L. (2000): Identification of genomic regions associated with stay green in sorghum by testing RILs in multiple environments. *Theor. Appl. Genet.* 100: 1225-1232.

- Tinker, N.A., Mather, D.E. (1995): MQTL: software for simplified composite interval mapping of QTL in multiple environments. Journal of Agricultural genomics vol 1: <http://www.ncgr.org/jag/papers95/paper295/indexp295.html>
- Utz, H.F., Melchinger, A.E. (1996): PLABQTL: A program for composite interval mapping of QTL. Journal of Agricultural Genomics 2: <http://www.ncgr.org/jag/papers96/paper196/indexp196.html>
- van Heusden, A.W., Koornneef, M., Voorrips, R.E., Bruggemann, W., Pet, G., Vrielink van Ginkel, R., Chen, X., Lindhout, P. (1999): Three QTLs from *Lycopersicon peruvianum* confer a high level of resistance to *Clavibacter michiganensis* ssp *michiganensis*. Theor.Appl.Genet. 99: 1068-1074.
- van Ooijen, J.W. (1992): Accuracy of mapping quantitative trait loci in autogamous species. Theor.Appl.Genet. 84: 803-811.
- Van Ooijen, J.W., Maliepaard, C., (1996): MapQTL (tm) version 3.0: Software for the calculation of QTL positions on genetic maps. CPRO-DLO, Wageningen
- Visscher, P.M., Thompson, R., Haley, C.S. (1996): Confidence intervals in QTL mapping by bootstrapping. Genetics 143: 1013-1020.
- Visscher, P., Whittaker, J.H., Jansen, R. (2000): Mapping multiple QTL of different effects: Comparison of a simple sequential testing strategy and multiple QTL mapping. Mol. Breeding 6: 11-24.
- Waldron, B.L., Moreno-Sevilla, B., Anderson, J.A., Stack, R.W., Froberg, R.C. (1999): RFLP mapping of QTL for fusarium head blight resistance in wheat. Crop Sci. 39: 805-811.
- Wang, D.L., Zhu, J., Li, Z.K., Paterson, A.H. (1999): Mapping QTLs with epistatic effects and QTLxenvironment interactions by mixed linear model approaches. Theor.Appl.Genet. 99: 1255-1264.
- Weller, J.I. (1986): Maximum likelihood techniques for the mapping and analysis of quantitative trait loci with the aid of genetic markers. Biometrics 42: 627-640.
- Whittaker, J.C., Thompson, R., Visscher, P.M. (1996): On the mapping of QTL by regression of phenotype on marker-type. Heredity 77: 23-32.
- Xu, W.W., Subudhi, P.K., Crasta, O.R., Rosenow, D.T., Mullet, J.E., Nguyen, H.T. (2000): Molecular mapping of QTLs conferring stay-green in grain sorghum (*Sorghum bicolor* L. Moench). Genome 43: 461-469.
- Yi, N.J., Xu, S.Z. (2000): Bayesian mapping of quantitative trait loci for complex binary traits. Genetics 155: 1391-1403.
- Young, N.D. (1999): A cautiously optimistic vision for marker-assisted breeding. Mol. Breeding 5: 505-510.
- Zeng, Z.B. (1993): Theoretical basis for separation of multiple linked gene effects in mapping quantitative trait loci. Proc. Natl.Acad.Sci.USA 90: 10972-10976.
- Zeng, Z.B. (1994): Precision of quantitative trait loci. Genetics 136: 1457-1468.

QTLS FOR ROOT GROWTH AND DROUGHT RESISTANCE IN RICE

Adam Price

Department of Plant and Soil Science, University of Aberdeen, Aberdeen AB24 3UU, UK

1. INTRODUCTION

Since the development of molecular markers first allowed the construction of saturated linkage maps, it was clear that quantitative trait loci (QTL) analysis could be usefully employed in analyzing the genetics of complex traits. By producing mapping populations based on crosses of parental varieties contrasting for the trait of interest, it should be possible to identify which parts of the genome can improve the trait, which parts influence component traits theoretically linked to main trait and approximately quantify the contribution of these component traits. Once achieved, the targeting of genomic regions for varietal improvement would be possible through marker assisted selection.

The application of this approach to the genetic dissection of drought resistance in rice (*Oryza sativa* L.) could be considered a challenge. The task is worthwhile considering the very substantial impact of drought on rice and the relatively limited progress previously made in improving the drought resistance of high yielding varieties (see Price and Courtois 1999; Fukai and Cooper 1995; Nguyen et al. 1997). However, it should be difficult for two main reasons. Firstly, the phenomenon of drought is complex in itself. To a plant, drought is an interaction between precipitation, evapo-transpiration, irradiation, soil physical properties (including the soil hydrological and strength properties), soil nutrient availability and competition with neighboring plants. Secondly, rice germplasm displays a diverse range of mechanisms of drought resistance, including mechanisms of drought escape (short duration), drought avoidance (e.g. deep rooting, osmotic adjustment) and drought tolerance (dehydration tolerance). Because the drought phenomenon and the mechanisms of drought resistance interact (i.e. the same mechanism of drought resistance will not work for all drought environments) it will naturally prove difficult to identify regions of the rice genome which contribute to resistance to drought in a broad range of drought environments.

The ultimate goal of breeders is to identify QTLs that can increase yield under drought or at least increase yield stability under drought. Since yield in the absence of drought is itself a complex trait influenced by many component traits, and since under drought, these traits will interact with the drought environment and the resistance mechanisms, the true complexity of breeding for increased yield under drought is grasped.

The QTL approach for studying drought resistance has been used in other cereal crops such as sorghum, maize and rice. The advantage of using rice is the large number of markers (RFLPs and microsatellite), the relatively small genome size compared to other major cereal (and hence the progress towards genome sequencing) and the syntenic relationship between rice and other monocotyledonous crops (see chapter 9 in this book). By conducting such a genetic study on drought resistance in rice, the ultimate goal is to obtain markers for genes which contribute to improved drought resistance, and identify which rice varieties are potential sources for those genes, for the use in a marker assisted breeding programme. To achieve this, answers to following questions are being sought;

- 1/ Which parts of the genome contribute to components of drought resistance?
- 2/ Which parts of the genome contribute to performance under drought?
- 3/ To what extent do the regions identified contribute in contrasting environments?
- 4/ How are these genetic components distributed in the rice germplasm?
- 5/ What are the relative merits of individual drought resistance mechanisms?
- 6/ What is the underlying physiological and molecular basis for the QTLs identified?

Given these objectives and the difficulties highlighted above, the logical approach was to produce several mapping populations and conduct relevant phenotyping and genotyping to identify the regions of the rice genome that contribute to drought related traits. By combining field evaluations of plant performance under drought with physiological experiments on the underlying mechanisms of drought resistance it should be possible to quantify the value of different drought resistance mechanisms. The practicality of generating molecular maps and the theory of QTL analysis has been covered in earlier chapters in this book. This chapter reviews the progress that is being made using a QTL approach to improve drought resistance in rice.

2. COMPLEXITY OF THE DROUGHT RESISTANCE IN RICE

Drought can mean different things to different people. To a breeder, however, it is best described as shortfall of water availability sufficient to cause loss in yield. The duration (days to weeks) and timing (seedling, vegetative or reproductive stage) of drought fundamentally alters the impact on yield as does the interaction between the plant climatic conditions other than rainfall (temperature, irradiation, wind) and soil (topography, physics, chemistry). In identifying a strategy to breed more drought resistant lines the target environment should be known.

The rice growing environments in which drought is a problem are different, particularly in terms of capacity for root growth. In rainfed lowland rice, the plant grows in a paddy field which is prone to water deficit. The puddling of the soil in preparation for sowing produces a hard-pan layer just below the soil surface which is difficult for roots to penetrate. In upland rice, the plants grow in unbunded fields that have no puddling so there is less likely to be a hard layer. Deep water rice, which is sown in a dry field before monsoon rains, can be prone to drought in the initial stages of its establishment. Within each culture type (rainfed lowland, upland, deep water), there will also be a range of climatic and soil conditions.

3. TRAITS AS TARGETS FOR IMPROVEMENT

There are many traits that contribute potentially to the drought resistance of rice (see reviews of (Fukai and Cooper 1995; Nguyen et al. 1997; Price and Courtois 1999). A brief summary is given below, with greater detail on root systems given their demonstrated importance in drought resistance and their contribution to other potential constraints such as nutrient deficiency.

3.1 The root system

Since a plant obtains its water and mineral requirements from its roots and the availability of these resources often imposes a limit to plant productivity, it is difficult to overstate the importance of roots to plant productivity. Root development is fundamentally involved in the response to many plant stresses, in particular drought and mineral deficiency (also, many toxic stresses are primarily experienced by roots, and their response is crucial to the whole plant's response). The possession of deep and thick root system which allows access to water deep in the soil profile is considered important in determining drought resistance in upland rice and considerable genetic variation exists (Ekanayake et al. 1985; Fukai and Cooper 1995; O'Toole 1982; Yoshida and Hasegawa 1982). This trait may be less important in rainfed lowland rice,

where hardpans may severely restrict root growth. Here, the ability to penetrate a hard layer is considered important and genetic variation in the ability to penetrate a layer of hard wax has been demonstrated (Yu et al. 1995). This trait may also be useful in upland rice where high penetration resistance may limit rooting depth and where soils will harden as they dry. It is important to note several points here. First, the penetration of roots through uniform hard layers like wax is probably achieved through the possession of large root diameter which resists buckling (Cook et al. 1997), while when the impedance is due to a coarse textured sandy or stony horizon, it may be that thin roots would penetrate more easily. Second, the investment of carbon in a deep root system may have a yield implication because of lost carbon allocation to the shoot. Third, early access to water stored in lower soil layers may be of limited use in some drought scenarios if the plant exhausts deep-water reserves before they are replenished or before the plant is harvested (i.e. the terminal drought).

It is important to appreciate that the root growth is profoundly influenced by the environment. Adverse conditions (chemical or physical) directly inhibit root growth (e.g. low water potential or high/low temperature). The shoot environment can also indirectly influence root growth either via carbon supply or signaling processes (e.g. light interception, water status, nutrient status). It has been suggested that plants respond to shifts in resource supply by allocating carbon to the organ involved in capturing the limited resource (Thornley 1972). When light is limiting, plants invest in shoot biomass. When nitrogen is limiting, they invest in root production. At the mechanistic level, theories implicating sucrose supply (Farrar 1992), hormonal action (Jackson 1993) or a combination of both (Van der Werf and Nagel 1996) has been advanced to explain this phenomenon. Responses to drought or temperature are probably more complex due to a multiplicity of physical or biochemical processes directly affected. At the genetic level, the response of roots to the environment is poorly understood because roots are intrinsically difficult to study, particularly in the natural environment.

3.2 Shoot-related traits of drought resistance

Leaf rolling, which is pronounced in rice, reduces both the loss of water through transpiration and the absorption of heat and light radiation. It is widely used as an indication of the degree of drought stress experienced by the plant, and hence has been selected against traditionally in breeding for drought resistance. However, leaf rolling in some varieties has been shown to behave quite differently. These varieties are slow to roll leaves when leaf water content declines (Dingkuhn et al. 1989; Price et al. 1997b; Turner et al. 1986a). In genotypes differing in their leaf rolling tendency, the use of leaf rolling in the field will therefore be of little value as an indication of overall drought resistance.

The ability to rapidly close stomata under water-deficit will reduce water loss. The sensitivity of stomata to leaf water status has been shown to vary in rice (Dingkuhn et al. 1989; Dingkuhn et al. 1991; Price et al. 1997b; Turner et al. 1986a; Turner et al. 1986b). In addition to reducing transpiration, however, both leaf rolling and stomatal closure reduce carbon assimilation. Therefore, genotypes with particularly sensitive leaf rolling or stomatal closure may do poorly in drought where water supply is only limited for a short time (i.e. intermittent drought). Relevant to this consideration is the report that rice varieties differ in their water use efficiency (WUE), the ratio of carbon gained to water used (Dingkuhn et al. 1991). Some results on other crop species have indicated that WUE can be either positively or negatively related to production under stress, which is thought to be largely dependent on the genotype's capacity to sustain transpiration (suggesting that WUE alone might be questionable as a selection criterion) (Blum et al. 1993). Analysis of WUE generally relies on measuring carbon isotope discrimination (Farquar et al. 1989), a technique which must be used very carefully (for example, if data are not to be confounded by differences in water availability which might exist between plants with different root architectures). The most recent development is to link carbon isotope discrimination to oxygen isotope discrimination, which gives a better ability to distinguish between carbon isotope discrimination due to stomatal resistance or photosynthetic activity (Barbour and Farquhar 2000). This new development may improve the cost-effective determination of water use efficiency in crops including rice.

Rice has low non-stomatal resistance to water loss (compared to other grass species) probably due to a relatively thin layer of epicuticular wax (O'Toole 1982). There is, however, genetic variation for epicuticular wax thickness (O'Toole and Cruz 1983) and it may have potential for improving drought resistance (Nguyen et al. 1997).

Two shoot related mechanisms of drought tolerance are osmotic adjustment and tolerance of tissue dehydration. Osmotic adjustment is the ability of a plant to accumulate solutes actively in cells, increasing cell osmotic pressure in response to water stress, which provides a degree of turgor maintenance, necessary for growth, and potentially increases water extraction. Tissue dehydration tolerance is the ability to retain biochemical activity with lowered cellular water content. Both traits have been shown to vary between rice varieties during a slow drought experiment (Lilley and Ludlow 1996). Plants with the ability to adjust osmotically or tolerate dehydration may delay leaf rolling, stomatal closure and maintain leaf expansion with little cost, which should promote resistance particularly in the terminal drought situation (Turner et al. 1982). Tripathy et al. (2000) reported the potential value of membrane stability as a component of drought tolerance in rice. These authors indicate that maintaining ion balance under tissue water deficit is important in drought resistance and show that genetic variation exists in rice. As with

several of the other traits identified above there is no evidence as yet of any correlation between membrane stability and yield.

In crops other than rice, especially sorghum, the stay green ability has been identified as a trait of drought resistance and is the target of QTL mapping (Xu et al. 2000). It reflects the retention of green leaf area under severe drought. A similar phenomenon occurs in rice, where it has been suggested that plants which are able to retain green leaf area are better able to recover after drought and give good yield (Fukai and Cooper 1995). However, in rice the degree of genetic variability in this trait is not known and, like leaf rolling, leaf drying (as the opposite of stay green is termed) is used as a method of assessing the degree of stress severity experienced by the plant. If genetic variability in green leaf retention exists in the germplasm being tested however, it should be used with caution.

Another target for improving the drought resistance of rice advocated by Horton (2000) is the better use of absorbed light through improved resistance to photoinhibition and an enhanced capacity for non-photochemical quenching. Each of the traits highlighted above are physiologically, biochemically and genetically complex in themselves and interact with each other. Efforts continue to be made towards identifying the most appropriate trait(s) to use in selection for drought resistance and the identification of the contributing QTLs.

4. THE QTL APPROACH APPLIED TO ROOT GROWTH AND DROUGHT RESISTANCE

The ultimate goal for breeders is to improve yield under somewhat predictable water limitation or to improve yield stability under the more realistic conditions of uncertain water inputs. Since yield under drought is genetically complex, interacts strongly with the environment and requires a large quantity of seed for assessment, it has not proved to be the main target of QTL-based research on drought resistance. Studies to date have followed two main approaches. The first is to screen populations for indicators of drought resistance in field trials. In these, plants are grown in the dry season and drought is imposed by withholding water. The second approach is to conduct laboratory or greenhouse experiments to analyze component traits of drought resistance, thereby reducing the variability in the environment which hampers their measurement in the field. The individual competent traits that have received the most attention are root morphology, root penetration ability and osmotic adjustment. Having received less attention is leaf rolling ability, stomatal sensitivity and membrane stability. There is certainly opportunity to study carbon isotope discrimination as an indicator of water use efficiency

(Dingkuhn et al. 1991), non-stomatal resistance (O'Toole and Cruz 1983) and possibly both stay-green ability and photosynthetic response to high light under stress (Horton 2000).

The objective in studying component traits is to identify genomic regions contributing to a trait that could improve drought resistance. In order to be of convincing value, the QTLs must be shown to contribute in a range of environmental conditions, at least those comparable to the target environment. Therefore, there has been and will continue to be an emphasis on conducting experiments in a range of conditions in order to assess QTL stability across environment. Once stable QTLs for component traits of drought resistance are identified, the next step will be to introduce this QTL into a near isogenic background in order to elucidate the underlying physiological and molecular nature of the QTL and to evaluate the contribution to yield in the target environment.

Another important consideration when searching for valuable QTLs is the distribution throughout the rice germplasm. The classic approach to QTL analysis would be to make a population based on parental line displaying the widest divergence in the trait. However, QTLs identified this way may be of limited value if the elite lines that are the target for genetic improvement by marker assisted selection already contain these QTLs. With this in mind, the use of more than one mapping population in identifying QTLs for drought resistance traits is crucially important.

5. PROGRESS IN MAPPING DROUGHT RESISTANCE AND COMPONENT TRAITS

In a recent review, Mackill et al. (1999) identified 7 genomic regions in which there is evidence from two mapping population that traits related to drought resistance occur. This is probably an underestimate in light of growing evidence of QTL maps for drought related traits.

5.1 Mapping populations

There are 5 populations that have been used to study drought resistance-related traits. Details of the parental varieties used in each of the crosses and the type of mapping populations are given in Table 1. The table also gives an abbreviation for each population which will be used throughout this chapter. The number of markers in the Az x IR population has increased from 135 since the first published use of this population for mapping root traits (Yadav et al. 1997). Details of the map for the CT x IR population used in Zhang et al. (1999) and Tripathy et al. (2000) have not yet been published. All populations have been used to identify QTLs for traits related to drought

resistance while some have been used to identify QTLs for performance under drought in field drought screens.

5.2 Assessment of root growth

Champoux et al. (1995) grew the Co x Mo population in 1 m long, 0.1 m diameter, soil-filled tubes for 38 days in a screenhouse in the Philippines. They measured maximum root length, root thickness, root to shoot ratio and root dry weight per tiller. Yadav et al. (1997) did something very similar with the Az x IR population, but had 0.2 m diameter tubes and analysed, in addition, deep root weight (weight below 0.3 m). Hememalini et al. (2000) assessed a subset of 56 of the Az x IR population in 0.75 m long x 0.18 m diameter tubes buried in the field in India. Price et al. (1997a; Price and Tomos 1997) analysed root length and thickness in 178 individuals in the Ba x Az F₂ population grown hydroponically for 28 days. The descendents of this population, the Ba x Az population have been tested in soil boxes and in thin chambers of soil (1 x 0.3 x 0.015 m) using two contrasting soil environments and evaluated root length, root thickness root/shoot ratio, and deep root weight (below 0.9 m) after 56 days (Price et al. 1999). Zhang et al. (1999) reported the identification of root growth QTLs in the CT x IR population. In all of the root morphology screens described above, there is likely to be little or no resistance to root growth due to soil physical properties. This is almost certainly in contrast to the majority of soils in which rice will be grown, in which soil penetration resistance is likely to be significant, highly variable (due to soil heterogeneity) and strongly influenced by soil water content.

Four populations (all but the Ba x Az F₂) have so far been screened for root penetration ability using different, and mostly slight, modifications of the petroleum wax layer method developed by Yu et al. (1995) (Price et al. 2000; Ray et al. 1996; Zhang et al. 1999; Zheng et al. 2000). In this method, a layer of hard wax is buried in soil pots and the number of roots which are able to grow through this layer is evaluated. Penetration ability is generally expressed as the number of roots penetrating the hard layer over the total number of roots. In the CT x IR population, root pulling force has been assessed (Zhang et al. 1999). This method involves measuring the amount of force required to pull up a plant from the soil and has been successfully used in the past as an integrated method of assessing the size of the root system (O'Toole and Soemartono 1981).

Table 1 Mapping populations used to locate QTLs associated with performance under drought or traits associated with drought resistance.

Population (abbreviation used in text)	Type of Population*	Pop. Size	No. of Markers	Drought resistance-related traits	Field drought screens	References
Azucena x IR64 (Az x IR)	DH	135	135-175	Root morphology, Root penetration, Osmotic adjustment	3 sites, 2 seasons	Yadav et al. 1997 Zheng et al. 2000 Courtois et al. 2000 Hemamalini et al. 2000
Bala x Azucena (Ba x Az F ₂)	F ₂	178	85	Root morphology, Leaf rolling, Stomatal conductance		Price and Tomos 1997 Price et al. 1997b
Bala x Azucena (Ba x Az)	RIL	205	135	Root morphology, Root penetration	2 sites, 2 seasons	Price et al. 2000 Price et al. 1999
Co39 x Moroberekan (Co x Mo)	RIL	203	127	Root morphology, Root penetration, Osmotic adjustment	1 site, 2 seasons	Champoux et al. 1995 Ray et al. 1996 Lilley et al. 1996
CT9993-5-10-1-M x IR62266-42-6-2 (CT x IR)	DH	154	315	Root morphology, Root penetration, Osmotic adjustment, Membrane stability	2 sites, 1 season	Zhang et al. 1999 Tripathy et al. 2000

* DH = double haploid, RIL = recombinant inbred lines

5.3 Assessment of other drought resistance-related traits

The hormone abscisic acid (ABA) is associated with molecular and physiological responses to drought in plants. Since ABA is synthesized in the root and shoot and under drought, the quantity of ABA in the leaves could be either an indication of an adaptation of drought resistance or an indication of stress. Quarrie et al. (1997) produced an F₂ population known to segregate for the degree to which ABA accumulated in detached and dehydrated leaves. QTLs for leaf ABA content were identified, but it is difficult to appreciate the implications for drought resistance.

The ability to osmotically adjust has been assessed using the methodology described by Lilley and Ludlow (1996) on slowly droughted pot plant in the

Co x Mo population (Lilley et al. 1996) and the CT x IR population (Zheng et al. 1999). Leaf rolling and the speed of stomatal closure has been assessed only in unreplicated leaf excision experiments on the Ba x Az F₂ population (Price et al. 1997b) while membrane stability has been mapped in the CT x IR population (Tripathy et al. 2000). This later report had 110 pot-grown lines exposed to a severe and nearly uniform stress (in terms of relative water content) and measured ion leakage from leaf discs.

5.4 Assessment of performance under drought

Four of the populations have been used to identify traits (other than yield) related to performance under drought. The drought has been applied by ceasing irrigation on plants grown in a dry season, natural drought period. The traits used are leaf rolling, leaf drying and relative water content, relative growth rate and drought score which are used as indicators of plant water status. As such, they are indicators of the degree of drought avoidance. Leaf rolling has traditionally been used as an indication of the degree of drought stress, since droughted rice plant roll their leaves and the trait is easily scored. It is potentially limited because there is known to be genetic variation in the degree of leaf rolling (in relation to plant water status) and a QTL for the trait has been located near *sd-1* on chromosome 1 (Price et al. 1997b). Leaf drying is another easily scored trait and used as an indication of water status. A more accurate assessment of plant water status is relative water content, but this is much more difficult to assess and is typically only used once. The relative growth of drought-affected plants (relative biomass increase during drought period) has also been used to indicate the degree of stress or the response to stress (Courtois et al. 2000). Champoux et al. (1995), using the Co x Mo population, assessed leaf rolling in one dry season screens in the Philippines in which drought was applied at 3 different growth stages, at the seedling stage (27 days after sowing), early vegetative stage (35 days after sowing) and late vegetative stage (48 days after sowing). Courtois et al. (2000) screened 105 lines of the Az x IR population in 2 seasons at one site and in one season in another. One site was the same as used by Champoux et al. (1995), the other in India. Price et al. (1999), using the Ba x Az population, conducted 2 dry season screens (on 176 and 110 lines) in the same field site in the Philippines, and used an additional site in West Africa (140 line) in one dry season screen. A further dry season screen in West Africa (110 lines) has subsequently been conducted (A Price personal communication). Hemamalini et al. (2000) grew 56 lines of the Az x IR population in sunken tubes and subjected them to a field stress in India. Zhang et al. (1999) evaluated all 154 lines of the CT x IR population during the dry season at Ubon in Thailand, irrigation terminating at 45 days after sowing. Visual drought score was assessed. A subset of 100 lines of this population was tested for drought avoidance at the reproductive stage in Israel in one season using 2 drying cycles, one at 97 days after sowing, the other at 116 days after sowing.

Performance under drought in this screen was indicated by yield of droughted plants relative to controls.

5.5 Mapping across populations

The molecular markers used in the mapping populations include RFLPs. RFLP probes are available on request from two sources, Cornell University (probes prefixed RG, RZ and CDO) and the Rice Genome Project (RGP), Japan (probes prefixed C, G, L, R, V). The mapping positions of Cornell probes can be deduced either by reference to the extensive map published by Causse et al. (1995) or by visiting the Rice Genes website (<http://genome.cornell.edu/rice>) while those of the RGP can be deduced by reference to the map published by Kurata et al. (1994) or by visiting the RGP website (<http://rgp.dna.affrc.go.jp>). Comparisons between RFLPs from the two sources is more difficult but can be achieved either by reference to maps which use combinations of both sets of probes (e.g. Lu et al. 1997, Price et al. 2000) or to the Oryzabase website (<http://shigen.lab.nig.ac.jp/rice/oryzabase>) which shows cross-alignment of both RFLP types. Cases where QTLs have been identified in more than one population are summarized in the sections below. This co-location of QTLs across populations gives confidence to breeders that they are valuable. However, it must be noted that in a few parts of the genome, markers of the two main types are not easy to align and this might introduce errors. It must also be noted that a QTL that is detected only in one population and/or one environment may be valuable if it reflects genes rare in the rice germplasm or if the environment in which it was identified is highly relevant to breeding objectives.

5.6 QTLs for drought avoidance

Genomic regions in which drought avoidance assessment reveals QTLs in two or more populations are summarized in Table 2. Only three regions, on chromosomes 1, 4 and 9 are revealed in three populations, but 10 QTLs are identified in 2 populations. Several important observations become clear when studying the results of individual publications. Firstly, there are a large number of regions for which there is evidence of an effect on drought avoidance. Secondly, in those experiments where more than one environment or more than one growth stage were used, there is clearly a different set of QTLs identified, indicating strong environment by genotype interaction. This indicates that some QTLs only contribute under some environments or at specific growth stages. Finally, different traits (leaf rolling, leaf drying, RWC and relative growth rate) do not always map to the same place, indicating different underlying genetic and physiological mechanisms and emphasizing that, for rice, there is no single trait that can, as yet, be used as a true reflection of drought avoidance. As Table 2 shows, there are regions that make good candidates for marker assisted selection and merit closer examination.

5.7 QTLs for root growth

Genomic regions in which assessment of root morphology or penetration ability reveals QTLs in two or more populations are summarized in Table 3. While 7 regions contain QTLs detected in 2 populations, 8 regions are detected in at least three including regions on chromosome 2 and 9 in which 4 populations contain root growth QTLs. Two comments are important here. Firstly, several regions appear to control many different aspects of root morphology or growth. For example, the region of chromosome 2 at 125 cM has been shown to effect root thickness, maximum root length, rooting depth and root penetration ability. This may be due to the clustering of genes influencing different components of root growth or the action of a single gene that has a pleiotropic effect on many aspects of root form. Secondly, even though regions can be shown to effect root morphology in several populations, it is not always the same trait that is detected in each population. For example at 10 cM on chromosome 3 is a region with a QTL for root penetration ability in one population, root pulling force in another and root thickness and root to shoot ratio in another. If we assume that, as discussed above, there is likely to be a gene or group of genes here that affect many aspects of root growth then why, for example, were root penetration QTLs not detected in all of these population. Three answers seem plausible. Firstly, each population may have a different allelic version of the gene that has different affects on root growth. Secondly, the gene (although the same) acts slight differently in the different genetic background (i.e. has different epistatic interactions). Thirdly, the ability to detect the QTL was determined by the exact environment in which the assessment was made since in no occasion has two populations been tested in exactly the same conditions.

5.8 Comparing drought avoidance and root growth QTLs

Figure 1 shows the relative position of the QTLs identified in Tables 2 and 3. There are 9 regions in which QTLs for drought avoidance detected in multiple populations coincide with QTLs for root growth identified in multiple populations. This may imply that nearly 70% of drought avoidance QTLs are related to root growth QTLs. Caution is required in making this assumption however, because the identification of QTLs for drought resistance-related traits other than root growth has revealed that some coincide with QTLs for root growth traits.

It would not be useful to most readers to describe each QTL highlighted in the Tables and the Figure. None the less, three regions of particular importance will be described in greater detail, thereby giving an indication of the nature of the co-location of QTLs across populations, traits and environments.

One of the most striking observations gained from comparing QTL mapping of root traits is that chromosome 9 has QTLs detected in all populations, at about 80 cM. In this region, Yadav et al. (1997) reported QTLs for maximum root length and deep root weight, Hememalini et al. (2000) reported QTLs for root thickness, Zheng et al. (2000) reported a QTL for root thickness and Champoux et al. (1995) reported QTLs for root thickness and root/shoot ratio. Largely unpublished data by A Price has revealed QTLs for root/shoot ratio, root thickness, deep root weight and maximum root length under both well-watered and water-limited conditions. There is evidence on 3 populations that this region affects drought avoidance. Courtois et al. (2000) reported QTLs for leaf rolling QTLs in all screens, while Champoux et al. (1995) reported leaf rolling QTLs in all three growth stages. In the Ba x Az population, a QTL for leaf rolling in only one screen has been reported in this region (Price et al. 1999).

The region of chromosome 1 at about 150 cM appears to affect drought avoidance and several drought resistance-related traits. This is the same place and may well be related to the semi-dwarfing gene *sd-1*. Here, QTLs for root morphology have been located in three populations. A QTL here affected all measured traits in Az x IR (Yadav et al. 1997) and several traits in Ba x Az when grown under well-watered conditions (Price et al. 1999 and unpublished data). In the Ba x Az population, QTLs for drought avoidance including leaf rolling were also located here (Price et al. 1999). The fact that leaf rolling QTLs are found here in this population is unsurprising since it has already been shown using a leaf excision test that there is a QTL affecting the response of leaf rolling to leaf water content in this population (Price et al. 1997b). A QTL for membrane stability has also been identified in this region in the CT x IR population (Tripathy et al. 2000), which suggests that there may be multiple drought resistance-related traits mapping to the same place due to pleiotropy or linkage.

A region of chromosome 2, at 125 cM has revealed QTLs for root penetration ability in 3 of the 4 populations in which RPI has been tested (Price et al. 2000; Zhang et al. 1999; Zheng et al. 2000). In addition, QTLs for maximum root length and root thickness have been mapped here in the Az x IR (Yadav et al. 1997) and Ba x Az (A Price unpublished data). In the Co x Mo population, QTLs for root length and thickness (Champoux et al. 1995) and root penetration (Ray et al. 1996) at marker *RG139* appears to be 20 cM away and may be unrelated. Surprisingly, there is not strong evidence for drought avoidance QTLs here.

Table 2 QTLs associated with performance under drought identified in more than one population. [Positions relative to the Rice Genome Project map at web site <http://rgp.dna.affrc.go.jp>, calculated by using reference maps of Causse et al. (1995), Kurata et al. (1994) and Lu et al. (1997) in addition to those referred to in the table. See text for explanation.]

Chromosome	Position (cM)	Population Used	Traits Identified*	Reference
1	75	Az x IR	LD, RWC, RGR	Courtois et al. 2000
		Ba x Az	LD, RWC	Price et al. 1999
		Co x Mo	LR x 2	Champoux et al. 1995
3	20	Az x IR	LR	Courtois et al. 2000
		Ba x Az	RWC	Price et al. 1999
3	~100	Az x IR	LR, RWC, RGR	Courtois et al. 2000
		Co x Mo	LR x 2	Champoux et al. 1995
4	30	Az x IR	DS	Hemamalini et al. 2000
		Ba x Az	RWC	Price et al. unpublished
		Co x Mo	LR x 3	Champoux et al. 1995
4	~110	Az x IR	LR x 3	Courtois et al. 2000
		Co x Mo	LR x 3	Champoux et al. 1995
5	85	Az x IR	LR x 3, LD x 2	Courtois et al. 2000
		Ba x Az	LR x 2, RWC	Price et al. unpublished
7	50	Az x IR	LR, RGR	Courtois et al. 2000
			LR	Hemamalini et al. 2000
		Ba x Az	LR	Price et al. 1999
7	80-115	Az x IR	LD, RGR	Courtois et al. 2000
		Co x Mo	LR x 2	Champoux et al. 1995
8	85-95	Ba x Az	RWC	Price et al. unpublished
		Co x Mo	LR x 3	Champoux et al. 1995
9	80	Az x IR	LR x 3	Courtois et al. 2000
		Ba x Az	LR	Price et al. 1999
		Co x Mo	LR x 3	Champoux et al. 1995
10	60	Az x IR	LD	Courtois et al. 2000
		Ba x Az	LR	Price et al. 1999
11	85	Az x IR	LD	Courtois et al. 2000
		Co x Mo	LR x 3	Champoux et al. 1995
12	110	Az x IR	DS	Hemamalini et al. 2000
		Ba x Az	LR	Price et al. 1999

* LR = Leaf rolling, LD = Leaf drying, RWC = relative water content, RGR = relative growth rate (drought relative to control), DS = drought score

6. PROGRESS IN MARKER ASSISTED SELECTION

Several of the genomic regions identified in the analysis described above have been chosen as targets for marker assisted selection. The reciprocal transfer of root growth QTLs on chromosomes 1, 2, 7 and 9 in the Az x IR population (Shen et al. 1999) has been completed (B. Courtois, personal communication). In that work, four segments of the Azucena genome have been transferred to the IR64 background, and four segments of the IR64 genome have been transferred in the Azucena background. Another project with greater focus on

applied breeding has transferred four root growth QTLs on chromosomes 2, 7, 9 and 11 from Azucena into a popular Eastern Indian upland variety, Kalinga III (A Price, J Witcombe, K Steele and B Courtois, personal communication). The residual heterozygosity in the Ba x Az population is being exploited to obtain near isogenic lines at 6 root growth and drought resistance QTLs. In this approach, individual recombinant inbred plants which are heterozygous for a QTL (and not heterozygous elsewhere) are selfed to produce plants which are homozygous for both parental types just at the QTL. It has yet to be verified if any of these near isogenic lines (Azucena x IR64, Azucena x Kalinga III or Bala x Azucena) do have contrasting root growth or drought resistance. Once it can be verified that the near isogenic QTL has affected either trait, the lines will be valuable tools for the further characterization of the QTLs responsible since these near isogenic lines can be used to study the pleiotropic effect of the underlying gene, investigate the underlying physiological processes involved and lead the way to gene tagging and gene identification/isolation.

7. QTL X ENVIRONMENT INTERACTION

The importance of QTL x environment interaction is difficult to overstate. Many of the traits related to drought resistance are known to be affected by the environment. The degree to which this occurs and the consequences for the value of identified QTLs for target environments must be understood. To illustrate, for a root growth QTL to be valuable in conferring drought resistance, it must be expressed. Yet root growth is known to be highly dependent on the environment in terms of rate of elongation growth and branching. Therefore, it is important to appreciate the degree to which a potentially valuable QTL is expressed in the target environment. This is why there has been much emphasis on conducting drought screens in multiple sites and years. The result is a clear indication that in some years or sites, some QTLs are not detected, suggesting that they may be of limited value. The same degree of multi-environment screening has not been done for roots or other drought resistance mechanisms and this must be a valuable endeavor for future research.

Table 3 QTLs associated with root morphology identified in more than one population.

Chromosome	Position (cM)	Population used	Traits identified*	Reference
1	150	Az x IR	RDW, DRW, MRL, RT	Yadav et al. 1997
		Ba x Az	DRW, DRN	Price et al. (unpublished)
		CT x IR.	RPF	Zhang et al. 1999
2	~35	Co x Mo	RT	Champoux et al. 1995
		CT x IR.	RT	Zhang et al. 1999
2	125	Az x IR	RT, MRL	Yadav et al. 1997
		Az x IR	RPI	Zheng et al. 2000
		Ba x Az	RPI	Price et al. 2000
		Ba x Az	DRW, MRL	Price et al. (unpublished)
		Co x Mo	RPI	Ray et al. 1996
		CT x IR	RPI, RD, RT	Zhang et al. 1999
3	10	Ba x Az	RPI	Price et al. 2000
		Co x Mo	RT, R/S	Champoux et al. 1995
		CT x IR	RPF	Zhang et al. 1999
3	~100	Ba x Az	RPI	Price et al. 2000a
		CT x IR	RPI, RPF	Zhang et al. 1999
4	30	Ba x Az	MRL	Price et al. (unpublished)
		CT x IR	RPI	Zhang et al. 1999
4	~100	Az x IR	MRL	Hemamalini et al. 2000
		Co x Mo	RPI	Ray et al. 1996
		CT x IR	RPI, RPF	Zhang et al. 1999
5	85	Az x IR	RT	Yadav et al. 1997
		Az x IR	RV	Hemamalini et al. 2000
		Ba x Az F ₂	MRL	Price et al. 1997a
		Ba x Az	RPI	Price et al. 2000
7	80-115	Az x IR	MRL, RDW, DRW	Yadav et al. 1997
		Az x IR	RPI	Zheng et al. 2000
		Ba x Az	MRL, R/S, RT	Price et al. (unpublished)
		Co x Mo	RT	Champoux et al. 1995
		CT x IR	RT	Zhang et al. 1999
8	85-95	Az x IR	RT	Yadav et al. 1997
		Az x IR	RT	Hemamalini et al. 2000
		Co x Mo	RT	Champoux et al. 1995
9	60	Az x IR	RDW, DRW	Yadav et al. 1997
		Ba x Az	DRW, RT, R/S	Price et al. (unpublished)
		Co x Mo	RT, RDW/T, R/S	Champoux et al. 1995
		CT x IR	RT	Zhang et al. 1999
9	80	Az x IR	RDW, MRL	Yadav et al. 1997
		Az x IR	RT	Hemamalini et al. 2000
		Az x IR	RT	Zheng et al. 2000
		Ba x Az	DRW, MRL, RT, R/S	Price et al. (unpublished)
		Co x Mo	DRW, MRL, RT, R/S	Champoux et al. 1995

* RDW = root dry weight, DRW = deep root weight, MRL = maximum root length, RT = root thickness, DRN = deep root number, RPF = root pulling force, RPI = root penetration index, RD = rooting depth, RV = root volume, R/S = root to shoot dry weight ratio, DRW/T = root dry weight per tiller.

Table 3 continued QTLs associated with root morphology identified in more than one population.

Chromosome	Position (cM)	Population used	Traits identified*	Reference
10	~30-40	Ba x Az	RPI	Price et al. 2000
		Co x Mo	RT, R/S	Champoux et al. 1995
11	85	Ba x Az F ₂	MRL	Price et al. 1997a
		Ba x Az	RPI	Price et al. 2000
		Co x Mo	RT, MRL, R/S	Champoux et al. 1995
		Co x Mo	RPI	Ray et al. 1996
		CT x IR	RPF	Zhang et al. 1999
12	110	Ba x Az F ₂	RV	Price et al. 1997a
		Ba x Az	MRL	Price et al. (unpublished)
		Co x Mo	DRW/T	Champoux et al. 1995

* RDW = root dry weight, DRW = deep root weight, MRL = maximum root length, RT = root thickness, DRN = deep root number, RPF = root pulling force, RPI = root penetration index, RD = rooting depth, RV = root volume, R/S = root to shoot dry weight ratio, DRW/T = root dry weight per tiller.

8. THE TARGET ENVIRONMENT AND THE TARGET VARIETIES

Throughout this review, comments have been made that the QTL identified for marker assisted selection should be suitable for the target environment. Unfortunately, that target environment is not well characterized. Ideally, for representative geographical regions long-term climatic data should be gathered in order to predict the likely drought occurrence (in terms of frequency, severity and timing) in order to match the drought resistance mechanism to the environment (possibly through the use of geographic information systems (GIS)). Equally important will be the evaluation of soil conditions that may affect moisture holding or limit the value of root-based drought resistance genes.

Likewise, there is as yet limited use of locally adapted germplasm for varietal improvement. Modified local varieties are more likely to be acceptable to farmers, and contain the required genes conferring adaptation to local biotic and abiotic environment. The degree to which the varieties which are the targets for improvement contain or lack QTLs for drought resistance needs to be studied. The approach of using AFLPs to produce DNA fingerprint linkage blocks as proposed by Zhu et al. (1999) offers some promise in this regard.

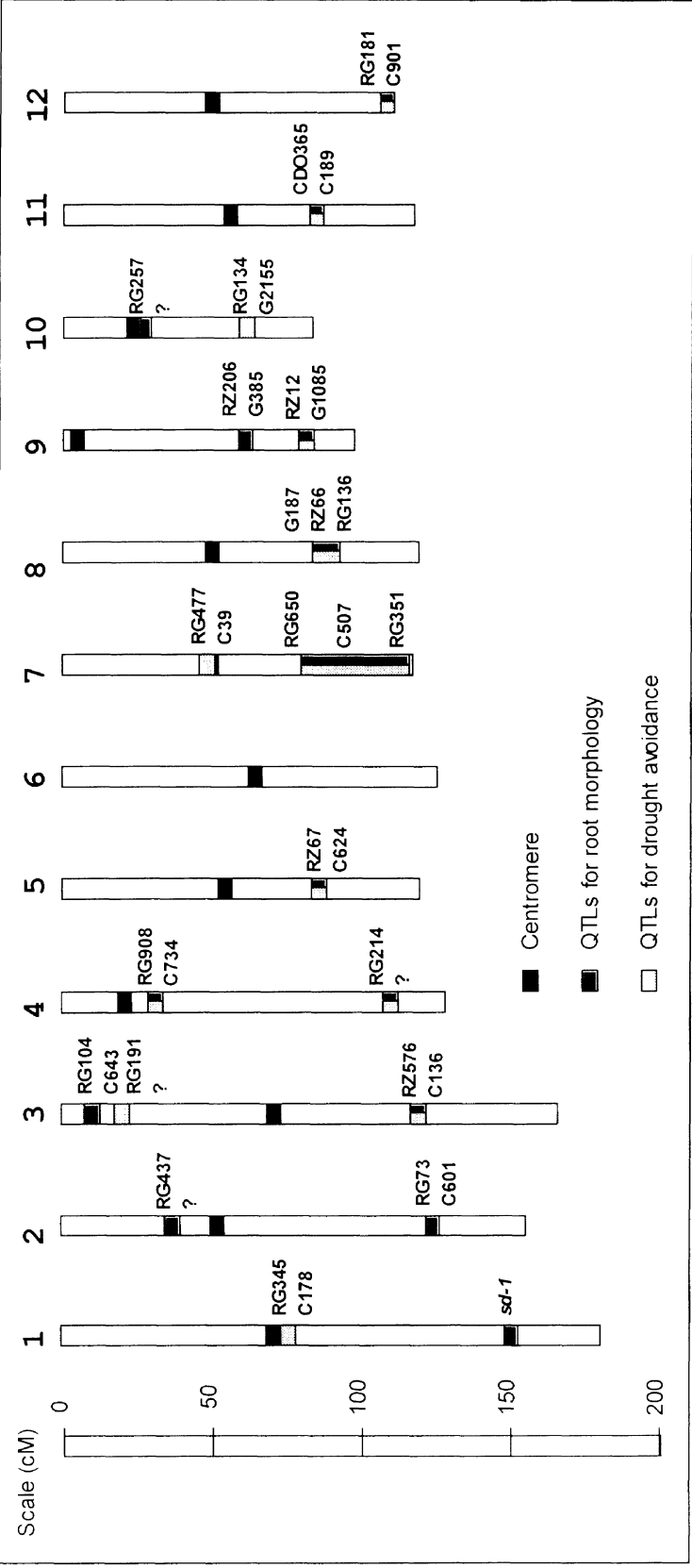


Figure 1. Molecular linkage map of the 12 chromosomes of rice showing evidence from more than one mapping population indicates the presence of QTLs associated with root growth or drought avoidance. Scale and orientation is according to the map published on the web by the Rice Genome Project, Japan.

Another consideration is the degree to which the close relatives of *Oryza sativa* might harbour positive genes for root growth or drought resistance, which have been lost in the adaptation of cultivated rice. The identification of novel QTLs for yield from a wild relative of rice, *O. rufipogon*, was reported by Xiao et al (1996). These authors demonstrated the principle that useful genes may be present in relatives of crop plants. A breeding programme at the West Africa Rice Development Association has had notable success in crossing *O. sativa* with *O. glaberrima* (a cultivated species grown only in Africa), thereby introducing drought resistance characteristics into an *O. sativa* background (see web site <http://www.cgiar.org/warda/research/nerica/nerica.htm>). QTL studies are currently being applied to these materials. There is probably much potential in exploiting this kind of genetic diversity in the relatives of *O. sativa*.

9. SUMMARY

The use of mapping populations of rice has enabled the identification of many QTLs associated with drought resistance in the field or traits theoretically linked to it. By careful comparison of data obtained from many populations and experiments, a pattern is beginning to emerge. Some genomic region are clearly candidates for marker assisted selection. These regions are also candidates for closer inspection via a near isogenic line approach, which should lead to greater understanding of the physiological and molecular mechanisms responsible. However, much remains to be done. The specific and detailed environmental conditions of the target regions needs to be more thoroughly evaluated. Many traits have not be thoroughly evaluated e.g. stomatal behavior, epicuticular wax, photoinhibition responses, water use efficiency, membrane stability. The relative merits of those traits well studied have not been evaluated. More attention needs to been given to the potential interaction between QTLs and the environment (this is especially true for root growth). The degree to which locally adapted varieties may lack the identified QTLs needs to be addressed. Close relatives of rice need to be exploited. With that said, the review given above indicates that there is great progress being made. This is very encouraging considering the very considerable difficulties associated with the challenge of improving root growth and drought resistance in rice.

10. REFERENCES

- Barbour MM, Farquhar GD 2000. Relative humidity- and ABA-induced variation in carbon and oxygen isotope ratios of cotton leaves. *Plant Cell Environ* 23:473-485
- Blum A 1993. Selection for sustained production in water-deficit environments. International Crop Science I. Crop Science Society of America, Madison, Wisconsin, USA

- Causse M, Fulton TM, Cho YG, Ahn SN, Chunwongse J, Wu K, Xiao J, Yu Z, Ronald PC, Harrington SB, Second GA, McCouch SR, Tanksley SD 1995. Saturated molecular map of the rice genome based on an inter specific backcross population. *Genetics* 138: 1251-1274
- Champoux MC, Wang G, Sarkarung S, Mackill DJ, O'Toole JC, Huang N, McCouch SR 1995. Locating genes associated with root morphology and drought avoidance in rice via linkage to molecular markers. *Theor Appl Genet* 90:969-981.
- Cook A, Marriott CA, Seel W, Mullins CE 1997. Does the uniform packing of sand in a cylinder provide a uniform penetration resistance? A method for screening plants for responses to soil mechanical impedance. *Plant Soil* 190: 279-287
- Courtois B, McLaren G, Sinha PK, Prasad K, Yadav R, Shen L 2000. Mapping QTLs associated with drought avoidance in upland rice. *Molecular Breeding* 6:55-66
- Dingkuhn M, Cruz RT, O'Toole JC, Dorffling K 1989. Net photosynthesis, water use efficiency, leaf water potential and leaf rolling as affected by water deficit in tropical upland rice. *Aust J Agri Res* 40: 1171-1181.
- Dingkuhn M, Farquhar GD, De Datta SK, O'Toole JC 1991. Discrimination of ^{13}C among upland rices having different water use efficiencies. *Aust J Agri Res* 42: 1123-1131.
- Ekanayake IJ, O'Toole JC, Garrity DP, Masajo TM 1985. Inheritance of root characters and their relations to drought resistance in rice. *Crop Sci* 25: 927-933
- Farquhar GD, Ehleringer JR, Hubick K 1989. Carbon isotope discrimination and photosynthesis. *Ann Rev Plant Physiol Plant Mol Biol* 40: 503-537
- Farrar, JF 1992. The whole plant: carbon partitioning during development. In: Carbon partitioning with and between organs (Eds. Pollock CJ, Farrar, JF and Gordon AJ), BIOS Scientific Publishers, pp163-180
- Fukai S, Cooper M 1995. Development of drought-resistant cultivars using physio-morphological traits in rice. *Field Crops Res* 40: 67-86
- Hemamalini GS, Shashidar HE, Hittalmani S 2000. Molecular marker assisted tagging of morphological and physiological traits under two contrasting moisture regimes at peak vegetative stage in rice (*Oryza sativa* L.). *Euphytica* 112:69-78
- Horton P 2000. Prospects for crop improvement through the genetic manipulation of photosynthesis: morphological and biochemical aspects of light capture. *J Ex Bot* 51:475-485
- Jackson, MB (1993). Are plant hormones involved in root to shoot communication? *Advances in Botanical Research* 19:104-187
- Kurata N, Nagamura Y, Yamamoto K, Harushima Y, Sue N, Wu J, Antonio BA, Shoura A, Shimizu T, Lin S-Y, Inoue T, Fukuda A, Shimano T, Kuboki Y, Toyama T, Miyamoto Y, Kiriwara T, Hayasaka K, Miyao A, Monna L, Zhong HS, Tamura Y, Wang ZX, Momma T, Umehara Y, Yano M, Sasaki T, Minobe Y 1994. A 300 kilobase interval genetic map of rice including 883 expressed sequences. *Nature Genet* 8: 365-372
- Lu C, Shen L, Tan Z, Xu Y, He P, Chen Y, Zhu L 1997. Comparative mapping of QTLs for agronomic traits of rice across environment by using a double-haploid population. *Theor Appl Genet* 94: 145-150
- Lilley JM, Ludlow MM 1996. Expression of osmotic adjustment and dehydration tolerance in diverse rice lines. *Field Crop Res* 48: 185-197
- Lilley JM, Ludlow MM, McCouch SR, O'Toole JC 1996. Locating QTL for osmotic adjustment and dehydration tolerance in rice. *J Ex Bot* 47: 1427-1436
- Mackill DJ, Nguyen HT, Zhang J 1999. Use of molecular markers in plant improvement programs for rainfed lowland rice. *Field Crop Res* 64:177-185
- Nguyen HT, Babu RC, Blum A 1997. Breeding for drought resistance in rice: Physiological and molecular genetics considerations. *Crop Sci* 37: 1426-1434
- O'Toole JC 1982. Adaptation of rice to drought-prone environments. In: Drought resistance in crops with the emphasis on rice, pp195-213. Manila: IRRI
- O'Toole JC, Cruz RT 1980. Response of leaf water potential, stomatal resistance, and leaf rolling to water stress. *Plant Physiol* 65: 428-432.
- O'Toole JC, Cruz RT 1983. Genotypic variation in epicuticular wax of rice. *Crop Sci* 23: 393-394

- O'Toole JC, Soemartono 1981. Evaluation of a simple technique for characterizing rice root systems in relation to drought resistance. *Euphytica* 30:283-290.
- Price AH, Courtois B 1999. Mapping QTLs associated with drought resistance in rice: progress, problems and prospects. *Plant Growth Reg* 29:123-133.
- Price A, Steele K, Townend J, Gorham G, Audebert A, Jones M, Courtois, B 1999. Mapping root and shoot traits in rice: experience in UK, IRRI and WARDA. In: Ito O, O'Toole J, Hardy B editors. Genetic improvement of rice for water-limited environments. Manila (Philippines): International Rice Research Institute. p 257-273.
- Price AH, Steele KA, Moore BJ, Barraclough PB, Clark LJ 2000. A combined RFLP and AFLP linkage map of upland rice (*Oryza sativa* L.) used to identify QTLs for root-penetration ability. *Theor Appl Genet* 100:49-56.
- Price AH, Tomos AD 1997. Genetic dissection of root growth in rice (*Oryza sativa* L.). II. Mapping quantitative trait loci using molecular markers. *Theor Appl Genet* 95:143-152.
- Price AH, Virk DS, Tomos AD 1997. Genetic dissection of root growth in rice (*Oryza sativa* L.) I: A hydroponic screen. *Theor Appl Genet* 95: 132-142
- Price AH, Young EM, Tomos AD 1997. Quantitative trait loci associated with stomatal conductance, leaf rolling and heading date mapped in upland rice (*Oryza sativa*). *New Phytol* 137: 83-91
- Quarrie SA, Laurie DA, Zhu J, Lebreton C, Semikhodskii A, Steed A, Witsenboer H, Calestani C 1997. QTL analysis to study the association between leaf size and abscisic acid accumulation in droughted rice leaves and comparisons across cereals. *Plant Mol Biol* 35: 155-165.
- Ray JD, Yu L, McCouch SR, Champoux MC, Wang G, Nguyen H 1996. Mapping quantitative trait loci associated with root penetration ability in rice (*Oryza sativa* L.). *Theor Appl Genet* 92: 627-636
- Shen L, Courtois B, McNally K, McCouch S, Li Z 1999. Developing near-isogenic lines of IR64 introgressed with QTLs for deeper and thicker roots through marker assisted selection. In: Ito O, O'Toole J, Hardy B editors. Genetic improvement of rice for water-limited environments. Manila (Philippines): International Rice Research Institute. pp 275-289.
- Singh K, Ishii T, Parco A, Huang N, Brar DS, Khush GS 1996. Centromere mapping and orientation of the molecular linkage map of rice (*Oryza sativa*). *Proc Nat Acad Sci USA* 93: 6163-6168
- Thornley JHM 1972. A balanced quantitative model for root:shoot ratios in vegetative plants. *Annal Bot* 36: 431-441.
- Tripathy JN, Zhang J, Nguyen TT, Nguyen HT 2000. QTLs for cell-membrane stability mapped in rice (*Oryza sativa* L.) under drought stress. *Theor Appl Genet* 100:1197-1202
- Turner NC. 1982. The role of shoot characteristics in drought resistance of crop plants. In: Drought resistance in crops with the emphasis on rice, pp115-134. Manila: IRRI
- Turner NC, O'Toole JC, Cruz RT, Namuco OS, Ahmad S. 1986a. Responses of seven diverse rice cultivars to water deficits I. Stress development, canopy temperature, leaf rolling and growth. *Field Crop Res* 13: 257-271.
- Turner NC, O'Toole JC, Cruz RT, Yamboa EB, Ahmad S, Namuco OS, Dingkuhn M 1986b. Responses of seven diverse rice cultivars to water deficits II. Osmotic adjustment, leaf elasticity, leaf extension, leaf death, stomatal conductance and photosynthesis. *Field Crop Res* 13: 273-286.
- Van der Werf A, Nagel O 1996. Carbon allocation to shoots and roots in relation nitrogen supply is mediated by cytokinins and sucrose: opinion. *Plant and Soil* 185:21-32
- Xiao J, Grandillo S, Ahn SN, McCouch SR, Tanksley SD, Li J, Yuan L 1996. Genes from wild rice improve yield. *Nature* 384:223-234
- Xu WW, Subudhi PK, Crasta OR, Rosenow DT, Mullet JE, Nguyen HT 2000. Molecular mapping of QTLs conferring stay-green in grain sorghum (*Sorghum bicolor* L. Moench). *Genome* 43:461-469
- Yadav R, Courtois B, Huang N, McLaren G 1997. Mapping genes controlling root morphology and root distribution on a double-haploid population of rice. *Theor Appl Genet* 94:619-632.

- Yoshida S, Hasegawa S 1982. The rice root system; its development and function. In: Drought resistance in crops with the emphasis on rice, pp83-96. Manila: IRRI
- Yu L, Ray JD, O'Toole JC, Nguyen HT 1995. Use of wax-petrolatum layers to simulate compacted soil for screening rice (*Oryza sativa* L.) root penetration ability. *Crop Sci* 35: 684-687
- Zhang J, Babu RC, Pantuwan G, Kamoshita A, Blum A, Sarkarung S, O'Toole JC, Nguyen HT 1999. Molecular dissection of drought tolerance in rice: from physio-morphological traits to field performance. In: Ito O, O'Toole J, Hardy B editors. Genetic improvement of rice for water-limited environments. Manila (Philippines): International Rice Research Institute. p 331-343.
- Zheng H, Babu R, Pathan M, Ali L, Huang N, Courtois B, Nguyen HT 2000. Quantitative trait loci for root-penetration ability and root thickness in rice: Comparison of genetic backgrounds. *Genome* 43:53-61.
- Zhu JH, Stephenson P, Laurie DA, Li W, Tang D, Gale MD 1999. Towards rice genome scanning by map-based AFLP fingerprinting *Mol Gen Genet* 261:184-195

QTL MAPPING OF FORAGE TRAITS

Thomas Lübberstedt

*Technical University Munich, Department of Agronomy and Plant Breeding, Alte Akademie
12, 85350 Freising, Germany*

1. INTRODUCTION

The ultimate goal of forage plant cultivation is to produce a maximum yield of metabolizable energy for livestock production. Although this seems to be a straightforward objective, forage agriculture is rather complex. A number of factors need to be considered, especially (i) the fed animal species, (ii) the forage plant species and cropping system, and (iii) the method(s) used for forage quality evaluation that determine the interaction between animal and feed.

Animals relevant for forage plant production such as cattle, sheep, goats, and deer are cud-chewing species, while nonruminating animals include pigs, horses, rabbits, and different bird species. The optimal feed is species-specific. Ruminants have a much better capability to handle fibrous carbohydrates compared to monogastrics and to convert poor quality protein and nonprotein nitrogen sources into human food (Van Soest 1974). Ruminants have varying preference for plants in pastures (Nösberger and Boberfeld 1986). The optimum diet of animals depends on the product, such as beef or milk in cattle.

A vast number of plants are used in animal nutrition. Forage plants can be classified according to (i) climate, (ii) cropping system, (iii) taxonomy, and (iv) on organs or plant parts used as fodder. Climatic classification depends mainly on the combination of humidity and temperature (McCloud and Bula 1974). Humidity differentiates regions into wet, humid, subhumid, semiarid, and arid humidity types, whereas temperature-based regions are classified as tropical, mesothermal, microthermal, taiga, tundra, and perpetual frost. Natural grasslands have developed in restricted precipitation under different climatic conditions (Trewartha et al. 1967), which, consequently, have induced adaption of diverse plant communities under these climates (Boberfeld 1986).

Forages are grown in a variety of cropping systems (Blaser et al. 1974, Nösberger and Boberfeld 1986). Methods of utilizing forages are associated with size of canopies, morphology, quality, stand longevity, and botanical composition. Forage utilization is a flexible combination of hay, silage, and grazing. Forage crops might be cultivated as mixtures or in monoculture as major or intermediate crop.

Although a large number of plant species are used as animal feed (Boberfeld 1986), the most important forage species belong to two families, the monocot grasses (Gramineae) and the dicot legumes (Leguminosae) (Table 1). Both families include more than 500 genera with annual, biennial, and perennial species. Together with Rhizobia, legumes fix atmospheric nitrogen. In most cropping systems, the above-ground parts of grass and legume forage plants are harvested before seed-set. Therefore, the quality of forage plants is mainly determined by their leaves and stems, especially by the composition of cell walls (Van Soest 1974).

The interaction between animals and forage described by quality parameters will be discussed in more detail below. However, it must be emphasized that frequently forages are fed along with cereal grains and vitamins. Furthermore, intake and digestion by animals depends on forage properties such as its dry matter concentration, particle size, and the ensiling process.

This review focuses on major forage grass and legume species relevant to industrialized countries in temperate climates, since biotechnology is currently employed in appreciable amount for these species (Table 1).

2. IMPORTANT FORAGE TRAITS

In most forage grasses and legumes, the above-ground parts are harvested before or during flowering. An exception is forage maize, which is harvested after seed-set. The major goals in forage production are (i) to maximize dry matter yield, and (ii) to achieve a high level of forage quality. Since both goals might be negatively correlated, the ultimate goal is to produce a maximum yield of metabolizable energy. In addition, the breeding objectives include resistance or tolerance to biotic and abiotic stress, maturity, mineral uptake capability, and competition with other species.

Dry matter yield can be easily determined. In contrast, quality evaluation is rather difficult. The direct approach to evaluate the quality of a given forage crop is the conduct of animal feeding trials to maximize the yield of the intended product such as milk or beef. Different parameters were developed for these traits such as Mega Joule “Net-Energy-Lactation“ ((MJ) NEL) (Groß 1979, Weißbach 1993) for milk production and “Kilo Starch-

Units“ (kStE) (Zimmer et al. 1980) for beef production and “Metabolizable Energy“ (ME) (Menke and Huss 1987), all reflecting the energy density (J/kg) of forage dry matter (Boberfeld 1986). Nevertheless, this direct approach of quality evaluation has a number of limitations. Animal trials are rather time-consuming, laborious, and consequently expensive. Hence, it is not possible to handle thousands of plant genotypes, as required in breeding programs. Furthermore, feeding trials depend on a number of additional factors, such as the animal species and genotypes employed, and the mode of feeding forage genotypes, impairing the generalization of the results. In consequence, a number of indirect biological, chemical, and physical methods for quality evaluation have been developed. Furthermore, prediction of the breeding value of forage plant genotypes based on molecular markers would be highly desirable.

Biological methods for quality evaluation can be subdivided into field, *in vitro*, and enzymatic methods. Field methods score the expected nutritive value of plant communities (Boberfeld 1986) based on the species composition. In case of forage maize, the proportion of ears in total dry matter has been used for quality evaluation (Zscheischler 1990). A widely used *in vitro* rumen digestion analysis was developed by Tilley and Terry (1963) using a two-step procedure - first rumen liquor and subsequently peptic hydrochloric acid to estimate the *in vitro* digestibility of organic matter. Another *in vitro* test employing rumen liquor determines gas production, protein and fat content to estimate NEL or StE (Menke and Steingass 1987). Enzymatic methods use cellulase together with peptic hydrochloric acid to estimate NEL and StE (Kirchgessner and Kellner 1981).

Since digestibility is mainly limited by poorly digestible cell wall components, chemical methods for forage quality evaluation focus on the breakdown and characterization of cell wall fractions within the organic matter. Using detergents Van Soest (1974) separated cell complexes into soluble cell content and insoluble “neutral detergent fibre” (NDF) representing mainly the cell wall fraction. By acidic detergents further fractionation into a lignin (“acid detergent lignin”: ADL) and cellulose fraction (ADF-ADL; ADF: “acid detergent fibre”) is possible. ADF values can be converted into NEL and StE estimates by convenient equations (Kirchgessner and Kellner 1981).

All above approaches are too laborious for routine quality evaluation of large numbers as required in plant breeding. Near-infrared reflectance spectroscopy (NIRS) (Norris et al. 1976) helps to overcome this limitation. By this method, large sample numbers can be investigated with low effort. Infrared spectra of ground materials (1400 to 2600 nm) can be employed to

estimate a number of quality parameters, if suitable calibrations exist to animal trials, biological or chemical methods.

3. CONVENTIONAL FORAGE CROP BREEDING

Biologically relevant aspects in forage crop breeding are the availability of self-incompatibility systems, interspecific and inter-generic hybridizations, polyploidy, and perennial nature of many species. Self-incompatibility, male sterility, and irregular chromosomal behaviour can be employed for controlled crosses but impair selfing and inbred line production (Johnson and Beyer 1974). Taxonomic relationships within families are favourable, since desired genes can be transferred through wide crosses. Furthermore, syntenic relationships across related species might accelerate identification of relevant genome regions (<http://ukcrop.net/comparative.html>). Autopolyploidy is of agronomic interest, if performance of polyploids is superior to that of diploids as in alfalfa. Autopolyploidy has also been used as first step to produce allopolyploids (*Festulolium*) and for the introgression of desired chromosomal segments into species (Boberfeld 1986). Perennial forage species compared to annuals require a long duration to evaluate new cultivars. This is one of the major reasons why methods for early evaluation of forage varieties would be beneficial.

Although forage crop breeding is as diverse as the number of forage species and traits, some properties of forage species can be generalized (Johnson and Beyer 1974). The majority of forage grasses and legumes are naturally allogamous (Smith 1944). For several species clonal propagation is possible but not economic. Therefore, the breeding strategy employed for most species is population breeding including production of synthetic varieties (Johnson and Beyer 1974, Becker 1993). In case of self fertility or when vegetative multiplication is possible, line or clonal breeding might be more appropriate (Johnson and Beyer 1974). Because of a large number of forage species and limited gain by breeding an individual crop, the establishment of hybrid breeding programs is in most cases not economic. An exception is the widely grown forage maize.

Given the prevalence of population and synthetic breeding of forage crops and the general three-step scheme of cultivar development, (i) generation of genetic variability, (ii) selection of parental genotypes, and (iii) testing of experimental populations or cultivars (Becker 1993), the following targets can be identified for forage crop breeding aided by modern technologies. For creating genetic variability, choice of base populations or parental genotypes is critical. Broadening of genetic variability and introgression of new traits within or across species would be most desirable.

During development of parental genotypes knowledge of the genome regions would help to increase allele frequencies in populations as well as during recurrent selection. For breeding of synthetics, genes that increase the general combining ability (GCA) for the desired traits would have high priority. Evaluation of experimental populations could be dramatically accelerated, if selection at the genomic or gene level would be possible.

4. APPLICATION OF MOLECULAR MARKERS IN FORAGE BREEDING

Molecular markers can be used at the genomic level or to link phenotype with genotype. The first application is used to determine genetic diversity within and between populations, pedigree relationships, and to build heterotic pools (Lee 1995). Association between genome regions represented by molecular markers and traits can be obtained from segregating populations. Segregation or bulked segregant analysis are conducted for simple inherited traits, and quantitative trait loci (QTL) analysis for complex traits (Lee 1995). These marker / genome region associations can be employed for marker assisted selection and marker-based gene isolation (Lee 1995).

Application of marker-based approaches requires the development of convenient genetic marker systems. After the first report on the concept of DNA-based molecular markers (Botstein et al. 1980), several marker systems were established for a large number of plant species. The major advantage of DNA-based markers is their virtually unlimited number allowing a dense coverage of whole genomes, their ease in application using any kind of plant tissue and their phenotypic stability. Some marker techniques such as RAPDs, ISSRs, and AFLPs can be immediately employed to any species, whereas development of RFLP, SSR, and SNP markers require preparatory work. Each marker system has specific advantages and disadvantages, reviewed extensively elsewhere (<http://www.plant.wageningen-ur.nl/about/Biodiversity/Cgn/research/molgen/default.asp>).

A literature search (CAB, AGRIS) in several forage species or genera did not detect any publication on molecular markers in many of them including *Phleum*, *Arrhenaterum*, *Alopecurus*, *Agrostis*, *Trisetum*, and *Melilotus*. Most marker-based studies were reported from *Medicago sativa*, *Lolium* / *Festuca*, and *Trifolium* ssp. (Table 1). These reports were mainly on marker and linkage map construction, and estimation of genetic

Table 1. Marker studies in forage crops.

Species /Genus	Marker type	Application	Reference
Grasses			
<i>Lolium perenne</i>	RFLP	Introgression of <i>Festuca mairei</i> chromosomes	Chen & Sleper 1999
	RFLP, RAPD, Isozyme	Linkage map	Hayward et al. 1994, 1998
	AFLP	Linkage map	Bert et al. 1999
	RAPD	Genotyping	Huff 1997
	AFLP	Genetic diversity	Roldan-Ruiz et al. 2000a
	AFLP	Genetic diversity	Roldan-Ruiz et al. 2000b
	AFLP	Variety description	Roldan-Ruiz et al. 2001
	SSR	Marker development	Forster et al. 2000a
	ISSR	Species-specific markers	Pasakinskiene et al. 2000
	AFLP	Physical (GISH) and genetic mapping	King et al. 2000
<i>Festuca</i> sp.	RFLP	Monosomic differentiation	Eizenga et al. 1998
	RFLP	Comparative mapping of meadow and tall fescue	Chen et al. 1998
<i>Poa pratensis</i>	RAPD	Core collection comparison	Johnson et al. 2000
<i>Dactylis glomerata</i>	AFLP	Major gene mapping	Barcaccia et al. 1998
	RFLP	Taxonomy (at chloroplast DNA level)	Sahuquillo and Lumaret 1999
Multispecies	STS	Cultivar differentiation	Lallemant et al. 1998
Multispecies	RFLP, RAPD	Linkage maps	http://ukcrop.net/perl/acc/search/ FoggDB
Legumes			
<i>Medicago sativa</i>	RFLP	Diploid inkage map	Brummer et al. 1993
	RFLP, RAPD	Diploid inkage map	Echt et al. 1994
	RFLP	Diploid inkage map	Tavoletti et al. 1996
	RFLP, RAPD	Diploid inkage map	Kiss et al. 1997
	RFLP, RAPD	Diploid inkage map	Kalo et al. 2000
	RFLP	Diploid inkage map	Brouwer and Osborn 1999
	RFLP	Cultivar identification	Labombarda et al. 2000
	AFLP	Fingerprinting	Barcaccia et al. 1999
	RFLP	Homozygosity test	Brouwer and Osborn 1997
	RFLP	Genetic diversity to establish synthetics	Kidwell et al. 1999
<i>Trifolium</i> sp.	RAPD	Taxonomy	Bullitta 1995
	RDNA	Taxonomy	Ansari et al. 1998
<i>Trifolium stoloniferum</i>	RAPD	Genetic diversity	Crawford et al. 1998
<i>Trifolium incarnatum</i>	RAPD, rDNA	Genetic diversity	Steiner et al. 1998
<i>Trifolium repens</i>	AFLP	Genetic diversity	Forster et al. 2000b

diversity. For most species, RAPDs or AFLPs were employed. RFLP markers were developed and frequently used in *Medicago sativa* and *Lolium perenne*, the most extensively investigated forage species at the molecular level. Since most studies were published in the past three years, an increasing number of reports may follow in the near future. This is substantiated by a number of home pages on the application of molecular markers in forage crops (Section 5)

5. QTL ANALYSES OF FORAGE TRAITS

Linkage maps are prerequisite for systematic QTL studies and available for *Medicago sativa*, *Festuca*, and *Lolium perenne*. These linkage maps have already been employed for mapping forage traits in *M. sativa* (Table 2) and a “stay green” gene in *L. perenne* and *Festuca* (Table 3). However, both reports on the mapping of forage QTL in forage crops are short presentations (Conference abstract, Database) and contain little additional information apart from QTL identification. In *M. sativa*, mapping of major genes and QTL has been published for a number of traits (Table 3). Since the basis for QTL mapping has been established for these three species and might be established for additional forage species, a number of QTL mapping activities are currently underway (Table 3).

Some agronomically important “non-forage” crop species can also be grown as forage crops, e.g., maize, canola, and rye (Boberfeld 1986). Although the molecular marker resources are much more elaborate for these “conditional” forage crops (<http://www.agron.missouri.edu/>, <http://ars-genome.cornell.edu/>, <http://wheat.pw.usda.gov/>), the major focus for mapping experiments in these species has been on grain yield, disease resistances, and morphological traits. At present, reports on mapping of forage yield and quality traits are available only for maize (Table 2). In contrast to a conference abstract by Dolstra et al. (1998), Lübberstedt et al. (1997a, b, 1998) published detailed studies in this area.

In maize, availability of molecular markers is not limiting for QTL studies (<http://www.agron.missouri.edu/>). Other critical questions for conducting QTL experiments are (i) choice of parent lines to establish a mapping population, (ii) the type and size of mapping populations, (iii) mode of trait evaluation, (iv) number of environments, (v) statistical data calculation, and (vi) experimental design (number of populations, testers). In the studies of Lübberstedt et al. (1997a, b, 1998), elite parent lines within the flint heterotic pool were selected for grain maize performance, since the mapping populations were also used in grain maize experiments (Schön et al. 1994, Melchinger et al. 1998). Usually parent lines extreme for the trait(s) of interest or giving at least a significant genetic variation in their progeny

are chosen with preference for elite compared to exotic germplasm to be close to the high performance level of breeding materials. Generally, populations can be developed with two or more genotypes per marker locus. In diploids, F_n populations (three genotypes per marker locus) allow estimation of dominance effects in contrast to BC, doubled haploid (DH) or recombinant inbred line populations. The major disadvantage of F_n families, employed in the studies of Lübberstedt et al. (1997a, b, 1998), is the genetic segregation within families increasing the experimental error.

Table 2. QTL analyses of forage traits.

Reference	Traits	Major results
<i>Medicago sativa</i>		
Shah et al. 1999	Forage yield, plant height, regrowth, growth habit, flowering time, leaf area, leaflet numbers	QTL identified for each trait, results not specified
<i>Zea mays</i>		
Dolstra et al. 1998	Cell wall digestibility (CWD), neutral detergent fibre (NDF), organic matter digestibility (OMD), amounts of different fermentation end products	Between 0-6 QTL were identified per trait, but results not specified
Lübberstedt et al. 1997a	Dry matter yield (DMY), Dry matter concentration (DMC), Starch yield (STY), Starch concentration (STC)	345 F_3 families and two testcross experiments; Heritabilities were intermediate to high; Between 7 and 16 QTL were detected for each tester-trait combination; Consistent QTL across testers were found for DMC and PHT, but not for DMY, STC, and STY
Lübberstedt et al. 1997b	In vitro digestible organic matter (IVDOM), Metabolizable energy yield (MEY), Metabolizable energy concentration (MEC), Protein yield (CPY), Protein concentration (CPC), Acid detergent fiber (ADF)	Same population and testers as above; Heritabilities were low to intermediate; Between 4 and 10 QTL were identified per tester-trait combination; Consistent QTL across testers were only found for CPY and CPC
Lübberstedt et al. 1998	DMY, DMC, IVDOM, CPC, PHT	Comparison across four populations using the same tester; Heritabilities were intermediate to high; Between 2 and 14 QTL were detected for each population-trait combination; Consistency across populations was greater for DMC, STC, and PHT compared to DMY, IVDOM and CPC

In several studies, mapping populations consisted of 100 to 200 families (Lee 1995). In this situation less QTL are detected compared to larger populations and estimates for gene effects are frequently inflated (Melchinger et al. 1998). This was corroborated by the studies of

Lübberstedt et al. (1997a, b, 1998) using one large “calibration” (345 F_3 families) and three small “validation” populations (113 to 140 F_n families). However, by comparing the large population with an independent small sample of the same cross, those QTL showing the largest gene effects in the large population were also detected in the smaller population (Lübberstedt et al. 1998). Hence, from a breeder’s perspective small populations might be sufficient for detecting the most relevant QTL.

Forage traits can be evaluated in whole plants or separately for the ear and stover fraction in forage maize. The stover seems to be mainly responsible for the genetic variation of forage quality in maize germplasm (Deinum and Struik, 1989). Nevertheless, Lübberstedt et al. (1997a, b, 1998) decided for whole-plant trait evaluations since this is easier to conduct and closer to agronomic practice, employing NIRS (Section 2). A larger number of environments increases experimental costs, but simultaneously trait heritabilities, which is an important prerequisite for QTL mapping (Utz and Melchinger 1994). Furthermore, the capability to estimate QTL x environment interactions is improved.

Advanced statistical procedures for QTL mapping such as composite interval mapping (Zeng 1994, Jansen 1994) have been developed and converted into software (e.g., PLABQTL: Utz and Melchinger 1996; QTL Cartographer, MapQTL: <http://linkage.rockefeller.edu/soft/list.html>) during the last decade and should be used for QTL studies to extract as much information as possible. QTL experiments can be confined to single populations. However, additional mapping populations or experiments (e.g., marker-assisted introgression) are required to validate detected QTL. In hybrid maize breeding, per se performance of inbreds is of minor interest compared to that of the hybrid performance. Therefore, four segregating populations were established within the flint heterotic pool and phenotypically evaluated after testcrossing to elite dent tester inbreds at the hybrid level (Lübberstedt et al. 1997a, b, 1998).

QTL studies provide (i) identification of genome regions and flanking markers affecting traits of interest, (ii) estimates of QTL positions, (iii) estimates of QTL gene effects including QTL x environment and epistatic interactions, (iv) understanding of trait correlations, and (v) evaluation of the predictive value of markers in marker-assisted selection. Lübberstedt et al. (1997a, b, 1998) detected between 2 and 16 QTL for each trait within each tester – population combination. Generally, the number of QTL detected increased with the population size and trait heritability. At the testcross level, only additive effects were estimated. QTL x environment and digenic epistatic interactions were rarely detected.

Trait correlations were reflected at the QTL level, e.g., for the negatively correlated traits of plant height and starch concentration all common QTL showed opposite additive effects. This indicates pleiotropic genes affecting the vegetative period. An extended vegetative period increases plant height and delays flowering, and therefore, decreases whole-plant starch concentration at harvest. Nevertheless, even in case of close trait correlations, trait specific QTL were detected. These are of interest to improve negatively correlated traits. Some traits, such as metabolizable energy (MEY) yield were calculated by multiplication of component traits (MEY: dry matter yield (DMY) * metabolizable energy concentration (MEC)). Almost all QTL of such complex characters (MEY) were attributable to QTL identified for the component traits. Hence, this procedure allows detection of those genome regions most relevant for MEY improvement (Section 2).

Table 3. QTL studies in forage crops and current QTL projects on forage traits.

Trait	Reference
Aluminium tolerance in alfalfa	Sledge et al. 1996
Seed proteins, leaf morphogenesis, soil nitrogen fixation in alfalfa	Kiss et al. 1997
Leaf and flower morphology in alfalfa	Brouwer and Osborn 1997
Lolium genome initiative	Forster 1999
<i>Medicago trunculata</i> as model species for legume Genomics	Cook 1999
Silage quality traits in maize	http://www.lars.bbsrc.ac.uk/cellbiol/devbio/mapmaiz.html
Silage quality traits in maize	http://www3.eureka.be/Home/projectdb/PrjFormFrame.asp?pr_id=2386
Stay green and other traits in <i>Lolium</i> and <i>Festuca</i> ; Comparative mapping	http://ukcrop.net/per/ace/search/FoggDB
Forage Genomics	http://www.noble.org/forgbiot/AR1999/Genomics/
QTL mapping of several traits in alfalfa in at least three North american research groups	http://www.naaic.org/Meetings/National/2000meeting/biotechreport.html
QTL mapping in alfalfa	http://www.public.iastate.edu/~Brummer/alfalfa.html

The predictive value of QTL was evaluated by comparing QTL results across testers within one population or across populations using the same tester. The three small validation populations had zero, one, or both parent lines in common with the large calibration population. Generally, the number of common QTL across populations increased with the genetic similarity of mapping populations. Almost all QTL detected in the small independent sample were also detected in the calibration population, both derived from the same cross, or could be explained by linked QTL with reversed gene effects. Some of these linked QTL were only identified in the small population due to higher resolution of its linkage map. For unrelated

mapping populations, about 70% of the detected QTL were specific to each population. However, consistency of QTL across populations as well as testers was highly trait-dependent (Table 2).

In general, anticipated and expected results of QTL studies as summarized above apply to any forage species with few exceptions. Mapped single copy markers (RFLPs, SSRs) were extensively used in maize (<http://www.agron.missouri.edu/>) and might also be useful for mapping of other forage species. QTL mapping using AFLPs or RAPDs might be easier, but transfer of information across populations is difficult without “core markers” (RFLPs, SSRs). Choice of parent genotypes to establish mapping populations is critical. Since QTL mapping is costly, a mapping population should be useful for a number of traits, and results should be applicable to breeding materials. While extensively investigated genotypes (inbreds) are available in maize, this might be not the case in at least some other forage species. Hence, pre-evaluation of potential parent genotypes or of potential mapping populations at the genotypic level might be reasonable. For some forage crops, such as alfalfa, commercial varieties are mainly tetraploid although diploids also exist. In this case, QTL mapping in diploids is much more straightforward. However, QTL detected at diploid level might not be functional at the polyploid level. A related question is the optimal mode of testing. Especially for breeding of synthetics, genome regions increasing GCA within a given synthetic are of highest priority. Hence, evaluation of testcross rather than per se performance (after selfing or cloning of mapped genotypes) will be preferable. Another topic relates to trait evaluation in monoculture or mixture with other species. The latter might be closer to agronomic practice but complicate QTL mapping. For all above mentioned questions it is important to know the correlations between experimentally favourable and agronomic or “genetic” (e.g., diploid - tetraploid) traits.

6. GENOMICS IN FORAGE CROP BREEDING

In contrast to marker-based approaches, Genomics aims at the characterization of genes by approaches like expression profiling (Bouchez and Höfte 1998). The ultimate goal is to deal with genes affecting a trait directly rather than with “indirect” markers linked to target trait genes. Genomics refers to application of global (in contrast to gene by gene) experimental approaches using genomic or cDNA (ESTs) sequence data to assess gene functions by parallel methods, such as expression profiling and screening for insertion mutants (Bouchez and Höfte 1998). Information gathered in one species might be transferred to another species (Comparative Genomics) by conserved map location of orthologous genes (synteny) in related species (Moore 1995), extraction of sequence databases and gene transformation. So far, scientific community focussed on a few model

species, such as the dicot *Arabidopsis thaliana*, rice, and maize (<http://www.mips.biochem.mpg.de/>, <http://www-iggi.bio.purdue.edu/>). However, these approaches are currently employed to a number of additional plant species including forages (<http://www.intl-pag.org>)

An important question in forage species is, whether the model species investigated are phylogenetically close enough to extract meaningful information by Comparative Genomics. This seems to be the case within the grass family (Moore 1995, <http://ukcrop.net/comparative.html>) including rice, maize, barley, wheat, rye, and sorghum (<http://www-iggi.bio.purdue.edu/>). Genes identified in these species might be directly used in forage grasses by gene transfer or for identification of orthologous genes to develop allele-specific markers.

The basis for Comparative Genomics is less advanced in forage legumes than grasses. However, Soybean is investigated in detail at the molecular level (<http://www.cbc.med.umu.edu/ResearchProjects/Soybean/Functional.html>). Another legume has been established as model species, the diploid *Medicago trunculata* (<http://www.bio.tamu.edu/medicago>). Within short time about 85.000 EST sequences (12 / 2000) were produced (http://ncbi.nlm.nih.gov/dbEST/dbEST_summary.html) and are presently employed in Functional Genomics projects (maize and rice had below 75.000 ESTs each). This Genomics project should provide much useful information at least for *M. sativa* but probably also for other forage legumes.

Forage crop breeding will profit from trait specific Genomics projects such as those exploring the molecular basis of cell wall composition (<http://plantgenome.sdsc.edu/projects.html>). Genes identified in these projects would be useful to manipulate forage quality either by gene transfer or by identification of orthologous genes.

7. CONCLUSIONS

Conventional forage crop breeding is complicated by their large number species and their performance evaluation. The selection process is not as clear-cut as in grain species, and the development of perennial species requires in forage crops a long duration for development and evaluation. The application of marker-assisted selection promises to speed up breeding of forage crops. Molecular approaches have been developed for major forages during the last decade, i.e., development of molecular markers and identification of some valuable genome regions by QTL mapping. Within the next decade, both application of molecular markers and Genomics should accelerate the identification of relevant genes or genome regions.

Furthermore, syntenic relationships among grasses or legumes should benefit the use of information across forage species. Thus, molecular tools will have an increasing impact on breeding of forage crop varieties.

ACKNOWLEDGEMENTS

I would like to thank both reviewers for their useful suggestions. Th. Lübberstedt is funded by a Heisenberg fellowship of the Deutsche Forschungsgemeinschaft (LU601/5).

8. REFERENCES

- Ansari, H.H., Reader, S.M., Badaeva, E., Ellison, N., Miller, T.E., Friebe, B. and Williams, W.M. (1998) Comparative FISH mapping of 18S-26S rDNA in six *Trifolium* species. Proc. V Plant & Animal Genome Conference, San Diego, CA.
- Barcaccia, G., Mazzucato, A., Albertini, E., Zethof, J., Gerats, A., Pezotti, M. and Falcinelli, M. (1998) Inheritance of parthenogenesis in *Poa pratensis* L.: auxin test and AFLP linkage analysis support monogenetic control, Theor. Appl. Genet. **97**, 74-82.
- Barcaccia, G., Albertini, E., Tavoletti, S., Falcinelli, M. and Veronesi, F. (1999) AFLP fingerprinting in *Medicago* ssp.: its development and application in linkage mapping, Plant Breed. **118**, 335-340.
- Becker, H. (1993) *Pflanzenzüchtung*, Ulmer Verlag, Stuttgart.
- Bert, P.F., Charmet, G., Sourdille, P., Hayward, M.D. Balfourier, F. (1999) A high-density molecular map for ryegrass (*Lolium perenne*) using AFLP markers, Theor. Appl. Genet. **99**, 445-452.
- Blaser, R.E., Wolf, D.D., and Bryant, H.T. (1974) Systems of grazing management, in: M.E. Heath, D.S. Metcalfe, and R.F. Barnes (eds.) *Forages (3rd edition)*, Iowa State University Press, Ames, Iowa, pp. 581-595.
- Boberfeld, W.O.v. (1986) *Grünlandlehre*, Ulmer Verlag, Stuttgart.
- Botstein, D., White, R.L., Skolnik, M. and Davis, R.W. (1980) Construction of a genetic linkage map in man using restriction fragment length polymorphisms, Am. J. Hum. Genet. **32** 314-331.
- Bouchez, D. and Höfte, H. (1998) Functional Genomics in plants, Plant Physiol. **118**, 725-732.
- Brouwer, D.J. and Osborn, T.C. (1997) Identification of RFLP markers linked to the unifoliate leaf, cauliflower head mutation of alfalfa, J. Hered. **88**, 150-152.
- Brouwer, D.J. and Osborn, T.C. (1997) Molecular marker analysis of the approach to homozygosity by selfing in diploid alfalfa, Crop Sci. **37**, 1326-1330.
- Brouwer, D.J. and Osborn, T.C. (1999) A molecular linkage map of tetraploid alfalfa (*Medicago sativa* L.) Theor. Appl. Genet. **99**, 1194-1200.
- Brummer, E.C., Bouton, J.H., Kochert, G. (1993) Development of a RFLP map in diploid alfalfa, Theor. Appl. Genet. **86**, 329-332.
- Bullitta, S. (1995) Development of RAPD markers of potential taxonomic use in the genus *Trifolium*, J. of Genet. And Breed. **49**, 313-318.
- Chen, C. and Sleper, D.A. (1999) FISH and RFLP-marker assisted introgression of *Festuca mairei* chromosomes into *Lolium perenne*, Crop Sci. **39**, 2676-1679.
- Chen, C., Sleper, D.A. And Johal, G.S. (1998) Comparative RFLP mapping of meadow and tall fescue. Theor. Appl. Genet. **97**, 255-260.
- Cook, D.R. (1999) *Medicago truncatula* as a model species for comparative and functional legume "Genomics", Proc. VII Plant & Animal Genome Conference, San Diego, CA, USA.

- Crawford, D.J., Esselman, E.J., Windus J.L. and Pabin, C.S. (1998) Genetic variation in running buffalo clover (*Trifolium stoloniferum*: Fabaceae) using random amplified polymorphic DNA markers (RAPDs). Proc. of the first Midwestern Rare plant Conf., St. Louis, Missouri, USA. Annals of the Missouri Botanical Garden, **85**, 81-89.
- Deinum, B., and Struik, P.C. (1989) Genetic variation in digestibility of forage maize and its estimation by near-infrared reflectance spectroscopy (NIRS). An Analysis, Euphytica **42**, 89-98.
- Dolstra, O., Van Loo, E.N. and Marvin, H.J.P. (1998) Genetic analysis of ruminal fermentation traits in forage maize. Proc. Plant & Animal Genome Conference VI, San Diego, CA.
- Echt, C.S., Kidwell, K.K., Knapp, S.J., Osborn, S.C., McCoy, T.J. (1994) Linkage mapping in diploid alfalfa (*Medicago sativa* L.), Genome **37**, 61-71.
- Eizenga, G.C., Schardl, C.L. Phillips, T.D. and Sleper, D.A. (1998) Differentiation of tall fescue monosomic lines using RFLP markers and double monosomic analysis, Crop Sci. **38**, 221-225.
- Forster, J.W. (1999) The international Lolium genome initiative, Proc. VII Plant & Animal Genome Conference, San Diego, CA, USA.
- Forster, J.W., Drayton, M., Dupal, M., Guthridge, K., Jones E. and Koelliker, R. (2000a) Development of efficient molecular marker systems for forage crop species, Proc. VIII Plant & Animal Genome Conference, San Diego, CA, USA.
- Forster, J.W., Koelliker, R., Jones, E., and Jahufer, Z. (2000b) Assessment of genetic diversity in white clover germplasm using AFLP markers, Proc. VIII Plant & Animal Genome Conference, San Diego, CA, USA.
- Groß, F. (1979) Nährstoffgehalt und Verdaulichkeit von Silomais. I.Mitt.: Bewertung von Silomais, Das wirtschaftseigene Futter **25**, 215-225.
- Hayward, M.D., McAdam, N.J., Jones, J.G., Evans, C., Evans, G.M., Forster, J.W., Ustin, A., Hossain, K.G., Quader, B., Stammers, M., Will, J.K. (1994) Genetic markers and the selection of quantitative traits in forage grasses, Euphytica **77**, 269-275.
- Hayward, M.D., Forster, J.W., Jones, J.G., Dolstra, O., Evans, C., McAdams, N.J., Hossain, K.G., Stammers, M., Will, J., Humphreys, M.O., Evans, G.M. 1998 Genetic analysis of *Lolium*. I. Identification of linkage groups and the establishment of a genetic map, Plant Breed. **117**, 451-455.
- Heath, M.E. (1974) Grassland agriculture in: M.E. Heath, D.S. Metcalfe, and R.F. Barnes (eds.) *Forages (3rd edition)*, Iowa State University Press, Ames, Iowa, pp. 13-20.
- Huff, D.R. (1997) Rapid characterization of heterogenous perennial ryegrass cultivars, Crop Sci. **37**, 557-564.
- Jansen, R.C., and Stam, P. (1994) High resolution of quantitative traits into multiple loci via interval mapping, Genetics **136**, 1447-1455.
- Johnson, I.J., and Beyer, E.H. (1974) Forage Crop Breeding, in: M.E. Heath, D.S. Metcalfe, and R.F. Barnes (eds.), *Forages (3rd edition)*, Iowa State University Press, Ames, Iowa, pp. 114-125.
- Johnson, R.C. Johnston, W.J., Nelson, M.C., Simon, C.J. and Golob, C.T. (2000) Comparing *Poa pratensis* core collections using agronomic and RAPD data. Proc. VIII Plant & Animal Genome Conference, San Diego, CA, USA.
- Kalo, P., Endre, G., Zimanyi, L., Csanadi, G. and Kiss, G.B. (2000) Construction and an improved linkage map of diploid alfalfa (*Medicago sativa*), Theor. Appl. Genet **100**, 641-657.
- Kidwell, K.K., Hartweck, L.M., Yandell, B.S., Crump, P.M., Brummer, J.E., Moutray, J. and Osborn, T.C. (1999) Forage yields of alfalfa populations derived from parents selected on the basis of molecular marker diversity, Crop Sci. **39**, 223-227.
- King, J., Thomas, H.M., Armstead, I.P., Morgan, G., Jones R.N., Kearsey, M.J., Donnison, I., Roberts, L., Harper, J.A. and King, I.P. (2000) Physical and genetic mapping of AFLPs in the forage grasses. Proc. VIII Plant & Animal Genome Conference, San Diego, CA, USA.

- Kirchgessner, M., and Kellner, R.J. (1981) Schätzung des energetischen Futterwertes von Grün- und Rauhfutter durch die Cellulasemethode, *Landw. Forschung* **34**, 276-281.
- Kiss, G.B., Kaló, P., Felföldi, K., Kiss, P. and Endre, G. (1997) Genetic linkage map of alfalfa (*Medicago sativa*) and its use to map seed protein genes as well as genes involved in leaf morphogenesis and soil nitrogen fixation. In: Biological fixation of nitrogen for ecology and sustainable agriculture. NATO ASI series G, Ecol Sci edn, vol 39. Springer, New York, pp 279-283.
- Labombarda, P., Pupilli, F. and Arconi, S. (2000) Optimal population size for RFLP-assisted cultivar identification in alfalfa (*Medicago sativa*), *Agronomie* **20**, 233-240.
- Lallemand, J., Lem, P., Kerlan, M.C. and Ghesquière, M. (1998) Development of STS markers for fodder grasses. Proc. VI plant & Animal Genome Conference, San Diego, CA.
- Lee, M. (1995) DNA markers and plant breeding programs, *Adv. Agron.* **55**, 265-344.
- Lübberstedt, Th., Melchinger, A.E., Schön, C.C., Utz, H.F., and Klein, D. (1997a) QTL mapping in testcrosses of European flint lines of maize: I. Comparison of different testers for forage yield traits *Crop Sci.* **37**, 921-931.
- Lübberstedt, Th., Melchinger, A.E., Klein, D., Degenhardt, H., and Paul, C. (1997b) QTL mapping in testcrosses of European flint lines of maize: II. Comparison of different testers for forage quality traits, *Crop Sci.* **37**, 1913-1922.
- Lübberstedt, Th., Melchinger, A.E., Fähr, S., Klein, D., Dally, A., Westhoff, P. (1998) QTL mapping in testcrosses of Flint lines of maize: III. Comparison across populations for forage traits, *Crop Sci.* **38**, 1278-1289.
- McCloud, D.E., and Bula, R.J. (1974) Climatic factors in forage production, in: M.E. Heath, D.S. Metcalfe, and R.F. Barnes (eds.) *Forages (3rd edition)*, Iowa State University Press, Ames, Iowa, pp. 372-383.
- Melchinger, A.E., Utz, H.F., and Schön, C.C. (1998) QTL mapping using different testers and independent population samples in maize reveals low power of QTL detection and large bias in estimates of QTL effects, *Genetics* **149**, 383-403.
- Menke, K.H., and Huss, W. (1987) *Tierernährung und Futtermittelkunde*, 3rd Ed. Ulmer Verlag, Stuttgart.
- Menke, K.H., and Steingäß, H. (1987) Schätzung des energetischen Futterwertes aus der in vitro mit Pansensaft bestimmten Gasbildung und der chemischen Analyse, II. Regressionsgleichungen, Übersichten zur Tierernährung **15**, 59-93
- Moore, G. (1995) Cereal genome evolution: pastoral pursuits with "lego" genomes. *Curr. Opin. Genet. and Development* **5**, 717-728.
- Norris, K.H., Barnes, R.F., Moore, J.E., Shenk, J.S. (1976) Predicting forage quality by infrared reflectance spectroscopy, *J. Anim. Sci.* **43**, 889-897.
- Nösberger, J., and Boberfeld, W.O.v. (1986) *Grundfutterproduktion*, Verlag Paul Parey, Hamburg.
- Pasakinskiene, I., Griffiths, C.M., Bettany A.J.E., Paplauskiene, V. and Humphreys, M.W. (2000) Anchored simple-sequence repeats as primers to generate species-specific DNA markers in *Lolium* and *Festuca* grasses, *Theor. Appl. Genet.* **100**, 384-390.
- Roldan-Ruiz, I., Dendauw, J., Van Bokstaele, E., Depicker, A., and De Loose, M. (2000a) AFLP markers reveal high polymorphic rates in ryegrasses, *Mol. Breeding* **6**, 125-134.
- Roldan-Ruiz, I., Calsyn, E., Gilliland, T.J., Coll, R., Van Eijk, M.J.T., and De Loose, M. (2000b) Estimating genetic conformity between related ryegrass (*Lolium*) varieties, II. AFLP characterization. *Mol. Breeding* **6**, 569-580.
- Roldan-Ruiz, I., Van Eeuwijk, F.A., Gilliland, T.J., Dubreuil, P., Dillmann, C., Lallemand, J., De Loose, M., and Baril, C.P. (2001) A comparative study of molecular and morphological methods of describing relationships between perennial ryegrass (*Lolium perenne* L.) varieties. *Theor. Appl. Genet.* (in press)
- Sahuquillo, E. and Lumaret, R. (1999) Chloroplast DNA variation in *Dactylis glomerata* L. taxa endemic to the Macaronesian islands, *Molecular Ecology* **8**, 1797-1803.

- Schön, C.C., Melchinger, A.E., Boppenmaier, J., Brunklaus-Jung, E., Herrmann, R.G., and Seitzer, J.F. (1994) RFLP mapping in maize: Quantitative trait loci affecting testcross performance of elite European flint lines, *Crop Sci.* **34**, 378-389.
- Shah, M.M., Luth, D., Brummer C.E., Council, C.L. and Kunz, R.C. (1999) Molecular mapping of QTLs for yield heterosis in tetraploid alfalfa. *Proc. VII Plant & Animal Genome Conference*, San Diego, CA.
- Smith, D.C. (1944) Pollination and seed formation in the grass, *J. Agr. Res.* **68**, 79-95.
- Steiner, J.J., Piccioni, E., Falcinelli, M., Liston, A. (1998) Germplasm diversity among cultivars and the NPGS clover collection, *Crop Sci.* **38**, 263-271.
- Tavoletti, S., Veronesi, F., Osborn, T.C. (1996) RFLP linkage map of an alfalfa mutant based on a F_1 population, *J. Hered.* **87**, 167-170.
- Tilley, J.M.A., and Terry, R.A. (1963) A two-stage technique for the *in vitro* digestion of forage crops, *J. Brit. Grassl. Soc.* **18**, 104-111.
- Trewartha, G.T., Robinson, A.H., and Hammond, E.H. (1967) *Elements of geography*, 5th edition, McGraw-Hill, New York.
- Utz, H.F., and Melchinger, A.E. (1994) Comparison of different approaches to interval mapping of quantitative trait loci, in: J. W. van Ooijen, and J. Jansen (eds.) *Proc. of the 9th Meeting of the EUCARPIA Section Biometrics in Plant Breeding*, Wageningen, The Netherlands, pp. 195-204.
- Utz, H.F., and Melchinger, A.E. (1996) PLABQTL: A program for composite interval mapping of QTL. *J. Quant. Trait Loci*. <http://probe.nalusda.gov:8000/otherdocs/jqtl>.
- Van Soest, P.J. (1974) Composition and nutritive value of forages, in: M.E. Heath, D.S. Metcalfe, and R.F. Barnes (eds.) *Forages (3rd edition)*, Iowa State University Press, Ames, Iowa, pp. 53-63.
- Weißbach, F. (1993) Bewerten wir die Qualität des Mais richtig ?, *Mais* **21**, 162-165.
- Zeng, Z.-B. (1994) Precision mapping of quantitative trait loci, *Genetics* **136**, 1457-1468.
- Zimmer, E., Theune, H.H. and Wermke, M. (1980) Estimation of nutritive value of silage maize by using chemical parameters and *in vitro* digestibility, in W.G. Pollmer and R.H. Phipps (eds.), *Improvement of quality traits of maize for grain and silage use*, Martinus Nijhoff Publishers, The Hague, Boston, London, pp. 447-465.
- Zscheischler, J. (1990) *Handbuch Mais*, DLG-Verlag, Frankfurt/Main.

9. CITED WEB-SITES

<http://ars-genome.cornell.edu/>
<http://linkage.rockefeller.edu/soft/list.html>
http://ncbi.nlm.nih.gov/dbEST/dbEST_summary.html
<http://plantgenome.sdsc.edu/projects.html>
<http://ukcrop.net/comparative.html>
<http://ukcrop.net/perl/ace/search/FoggDB>
<http://wheat.pw.usda.gov/>
<http://www.agron.missouri.edu/>
<http://www.bio.tamu.edu/medicago>
<http://www.cbc.med.umu.edu/ResearchProjects/Soybean/Functional.html>
<http://www.intl-pag.org>
<http://www.lars.bbsrc.ac.uk/cellbiol/devbio/mapmaiz.html>
<http://www.mips.biochem.mpg.de/>
<http://www.naaic.org/Meetings/National/2000meeting/biotechreport.html>
<http://www.plant.wageningen-ur.nl/about/Biodiversity/Cgn/research/molgen/default.asp>
<http://www.public.iastate.edu/~Brummer/alfalfa.html>
http://www3.eureka.be/Home/projectdb/PrjFormFrame.asp?pr_id=2386
<http://www-iggi.bio.purdue.edu/>

QTLS IN BREAD MAKING QUALITY : A REVIEW

G. Charmet, C. Groos

INRA, plant breeding and plant health unit, 234 av. du Brézet, F63039 Clermont-Ferrand cedex, France

INTRODUCTION

Bread wheat is one of the most widely cultivated crop around the world, including all temperate zones and altitude regions in the tropics. 583 millions tons have been produced in 1999 from 215 M hectares, the most contributing countries being China (117 Mt), India (70 Mt), USA (62 Mt) and France (37 Mt). Out of these nearly 600 Mt available each year, about 80% is used for human food. Among the huge diversity of wheat derived food products, bread is probably the major one, both for the proportion of wheat which is transformed into bread and for the cultural, even religious link it has for many people. There is a vast array of different types of breads around the world: white and pan breads, including buns, hearth breads, such as baguette, flat breads, consumed in North Africa, the Middle East and India, steamed breads most popular in China (Morris 1998)... Obviously these different products use different processes, and would require different quality of the raw product, wheat. However they all share some common features: they are all cooked products, as cooking is necessary improve the digestibility of flour components, and they are all derived from a dough, which is derived from a worked mix of flour and water (plus some additives like salt), fermented by yeast. Therefore the “quality” of bread (shape, aspect, flavour...), would depend not only on flour composition, but also on the quantity of water, the mixing process, and the complex interactions between flour components during mixing in an hydrated environment. The more complex a trait, the more many genes are expected to be involved in its determinism, and therefore the effect of each will account for only a small part of the variation. The consequence will be more difficulty in detecting and estimating the effects of these genes when working on the final trait, and the necessity to work out its components or more heritable related traits.

Quantitative trait loci have been first defined by Sax (1923), who found an association between coat colour, a trait which was governed by mendelian genes, and grain weight in bean. Although some attempts have been made to locate QTLs using isozyme loci, such studies remained scarce until the 1980s and the advent of molecular techniques such as RFLP, RAPD and AFLP. Indeed these methods have made available a huge number of genetic markers, which allowed the construction of saturated linkage maps in most crop species. The availability of such dense genetic maps has opened the way to systematic search of QTLs for many traits of interest. Obviously such costly genome exploration have been first focused on traits such as plant height or flowering date, because of their ease of measurement. Later on, molecular techniques became available to plant-breeders, and more complex traits, like yield or quality parameters, have been tentatively dissected into QTLs.

Bread making ability of wheat is known to be influenced by both genetical and environmental factors, thus having a low heritability and being difficult to select for (Oury *et al.* 1999). Almost all components of flour are supposed to play a role in bread-making dough properties : starch, pentosans, soluble sugars, proteins (different categories), lipids (free polar or associated to other components), enzymes.... Although the genetic determinism of some of these components is known (Morris 1998), there are still many more genes, involved either directly in the biosynthesis pathways, or indirectly as regulatory genes, which remain to be discovered. As a first approach, without *a priori* on putative candidate genes, QTL analysis may be useful in identifying the region of the genomes likely to bear the functional genes.

This paper aims to summarise results from the published QTL analyses of bread-making related traits.

BIBLIOGRAPHIC SURVEY

In order to get an extensive overview of world's research results up to end of year 2000, we did a bibliographic survey of CAB and current content databases using a range of key-words to make us fairly sure that we were able to gather most published results on the topic. Quite surprisingly references are not so numerous, although we must admit that we might have missed some. We have chosen to summarise the results into three tables, according to the traits which were under study. These tables are partly redundant, since some studies were actually multitrait. Thus a given reference may be found in several tables.

Table 1: QTLs for bread making scores and dough rheological tests in wheat

Trait	Cross	Progeny	MAP	Nb env	QTLs	Total or cumul R ²	Reference
Pelschenke and alveograph	Armada x F.A. Rex x Corin	243 RILs 183 RILs	Storage proteins Storage proteins	1	HMW, LMW Glu and Gli have effects	68 % 51 %	Khelifi and Branlard 1992
SDS and mixograms	Klein 32 x CS	96 F ₃ -SSD	15 storage proteins loci		IBs: XgliB1 and XgluB3	11-15%	Manifesto <i>et al.</i> 1998
W alveograph	Courtot x CS	187 DH	350 RFLP and SSRs		1A ^L , 3B, 5D ^S	39%	Perretant <i>et al.</i> 1999
Gluten strength	T. durum: Jennah Khelifa x Cham1	110 RILs	319 markers	13	1A, 1B (XgliB3), 4B, 6B	35%	Elouafi <i>et al.</i> 2000
Alveograph parameters	Foma (T. aest.) x Oberkulmer (spelt)	204 RILs	187 markers	4	10 QTLs, for W (major on 5A and 5D : 15%), 7 for P/L	39% (W) 38% (P/L)	Zanetti <i>et al.</i> 2001
Bread volume	CS* x Cheyenne 1A, 1B, 1D	50 RSL	RFLPs and storage proteins		1 on 1A, 2 on 1D	30%	Rousset <i>et al.</i> 2001

CS: Chinese Spring

RSL: Recombinant Substitution Lines

FA: Florence Aureore

SSD: Single Seed Descent
DH: Doubled Haploid lines

Nb env: Number of environments

RILs: Recombinant Inbred Lines

QTL for bread-making scores and rheological tests

By the end of year 2000, there are very few reports of studies of bread parameters in segregating populations (Table 1).

The only published one is that of Rousset *et al.* (2001), who explored homeologous group 1 using recombinant substitution lines and found QTLs for a range of parameters, including loaf volume, on chromosomes 1A and 1D, which may be explained by the storage protein loci. The other studies reports results of small-scale tests aimed to predict bread-making value. It is known however that such indirect tests show varying level of correlation to direct bread-making evaluation (Kieffer *et al.* 1998, Oury *et al.* 1999). From our experience, the correlation between indirect tests and bread-making score appears to be quite high ($r^2 > 0.5$). Thus the reliability of prediction tests may be higher in a given segregating population than the poor value reported in a broad base breeder's material.

The two main tools for evaluating rheological properties of dough on small flour quantities are mixograph, which measures dough parameters during mixing, and alveograph, which rather estimates dough strength and extensibility. QTLs for mixograph parameters have been studied by Manifesto *et al.* (1998) on a segregating population. However they did not scan the whole genome, as they only mapped 15 polymorphic storage protein loci. Similarly Khelifi and Branlard (1992) have studied the effect of storage proteins (HMW and LMW glutenins and Gliadins) on Pelschenke swelling tests and alveograph parameters in two cross progenies, and reported quite high coefficient of determination, which were consistent with previous results of Carillo *et al.* (1990). Perretant *et al.* (2000) carried out alveograph tests on a population of 187 doubled-haploid lines between cv Courtot (good bread-making quality) and Chinese Spring (poor). Zanetti *et al.* (2001) reported a similar study in their wheat by spelt population. They found 10 QTLs for dough strength, 4 for elasticity index and 7 for conformation ratio P:L. We can also consider results from Elouafi *et al.* (2000), who studied gluten strength on durum wheat.

From these results, it appears that chromosomes of homoeologous group 1, especially 1A and 1B, are more often mentioned as bearing QTLs for bread-making traits. However, this result is partly biased, since three of the studies mentioned in Table 1 considered either genetic material which was only segregating for these group 1 chromosome and used only storage protein loci for establishing linkages with QTLs. Among the other linkage groups, only chromosome 3B was found consistently in two genetically unrelated materials. The other chromosomes were only mentioned in a single study.

Table 2: QTLs for protein content and protein quality characterisation

Trait	Cross	Progeny	MAP	Nb env.	QTLs	Total or cumul R ²	Reference
Protein content	Wheat x <i>T. tauschii</i>	147 BC2	34 markers	6	2 not mapped	29%	Fritz <i>et al.</i> 1995
Protein content	Barley: Blenheim x Kim	99 DH	99 markers	1	8: 1H ^S , 1H ^L , 2H ^L , 5H ^S , 6H, 7H ^S , 7H ^L	NA	Bezant <i>et al.</i> 1997
Protein content	Langdon (durum) x L. 6B (<i>T. dicoccoides</i>)	85 RCILs			1QTL on 6B	66%	Joppa <i>et al.</i> 1997
Protein content	PH132 x WL711	100 RILs	57 markers, BSA	1	1 on 2D ^L	19%	Prasad <i>et al.</i> 1999
Protein content	Messapia (durum) x <i>T. dicoccoides</i>	65 RILs	259 markers	6	8 on: 4A ^L , 4B ^S , 5A ^L , 6A ^S , 6B ^L , 7A ^L , 7B ^S	49-56%	Blanco <i>et al.</i> 1996, 1998.
Sedimentation volume	Messapia (durum) x <i>T. dicoccoides</i>	65 RILs	259 markers	6	7: 1A ^L (XgluA1) 1B ^S (XgliB1/gluB3), 3A ^S , 3B ^L , 5A ^L , 6A ^L , 7B ^S	36-64%	Blanco <i>et al.</i> 1996, 1998.
Protein content	Forna (<i>T. aest.</i>) x Oberkulmer (spelt)	226 RILs	231 markers	3	6 QTLs, one major on 5A (25%)	49%	Zanetti <i>et al.</i> 1999, 2001
Sedimentation volume	Forna (<i>T. aest.</i>) x Oberkulmer (spelt)	226 RILs	231 markers	3	7: 1A ^S (XgluA3), 2A, 5A common with PC	45%	Zanetti <i>et al.</i> 1999, 2001
Protein content	Courtiot x CS	187 DH	350 RFLP and SSRs	3	1B ^L , 6A ^S	23%	Perretant <i>et al.</i> 1999
Protein content	Renan x Réctal	194 RILs	330 markers	6	2A, 3A, 4D, 7D	12-30%	Groos <i>et al.</i> submitted

SSD: Single Seed Descent RILs: Recombinant Inbred Lines DH: Doubled Haploid lines BSA: Bulk Segregant Analysis
 BC: Back Cross RCILs: Recombinant Inbred Chromosome Lines Nb env. : number of environments

QTL for protein content and protein quality in cereals

In Table 2, we summarised results from QTL analysis for grain protein content in bread wheat, durum wheat, or crosses between them and some wild related species and one study in barley, since a high level of syntenic relationships has been shown among these species.

Indeed grain protein content per se is known to be correlated with bread-making ability, at least with loaf volume of pan breads (Bushuk 1998). Thus crude protein content can be a goal in breeding for bread-making quality. Results gathered in Table 2 show that almost all linkage groups, except group 3, bear QTLs for grain protein content in durum wheat, spelt wheat or barley. Surprisingly enough, those chromosomes that harbour the storage protein genes are not the only ones involved, as far as the quantitative variations of these proteins are concerned. Obviously, not only the crude grain protein content, but also protein composition, i.e. the relative proportions of insoluble glutenins, soluble glutenins and gliadins are of primary importance for predicting gluten formation and dough properties. We are currently carrying out QTL studies for these relative ratios in the Recital x Renan progeny cited in Groos *et al.* (*in press*). Alternatively, sedimentation tests, such as SDS or Zeleny tests, provide an indirect estimate of the aggregation status of grain protein, particularly when corrected from grain protein content (specific Zeleny). Some studies report QTLs for these quick tests of protein quality. In the two studies (Blanco *et al.* 1998 on durum wheat and Zanetti *et al.* 1999 in a bread wheat x spelt progeny), quite a large number of QTLs have been found, almost in all homoeologous linkage groups, which altogether explained up to 64% of phenotypic variation. More evidently than for grain protein content, some of these QTLs for quantitative variation co-locate with the genes encoding either gliadins or LMW glutenins (1B^S) or HMW glutenins (1A^L), providing arguments for a direct role of protein structure in sedimentation volume.

QTL for other grain components and quality related traits

Obviously, proteins are not the only component of cereal grains, even not the most abundant. Although gluten proteins are supposed to play a major role in bread-making (Payne 1987), the influence of other grain features cannot be discarded. Results of the most significant QTL studies concerning these traits are presented in Table 3.

Table 3: QTLs for other grain components in cereals:

Trait	Cross	Progeny	MAP	Nb env.	QTLs	Total or cumul R ²	Reference
Pre-harvest sprouting (PHS)	NY18 x Clark's C NY18x NY10	78 F5 RILs 138 RILs	195 RFLP	6 7	4: 1A ^S , 2 ^S , 2 ^L , 4: 3B ^L , 4A ^L , 5D ^L , 6B ^L	37% 29%	Anderson <i>et al.</i> 1993
Hadberg's Falling number	Forna (T. aest.) x Oberkulmer (spelt)	226 RILs	231 markers	3	8: on 5A(Q), 6D ^L , 7B ^L (amy 1-3)	73%	Zanetti <i>et al.</i> 1999
PHS susceptibility	Renan x Réctal	194 RILs	330 markers	2	3A, 3B, 3D, 5A	38%	Groos <i>et al.</i> 2001- <i>In press</i>
Grain hardness	Wheat x T tauschii	147 BC2	34 markers	6	1 not mapped	9%	Fritz <i>et al.</i> 1995
Grain hardness	Synthetic x Opata	114 RILs	ITMI map (>1000 markers)	1	1 major on 5D ^S , 4:2A ^L , 2D ^L , 5B ^L , 6D ^S	70%	Sourdille <i>et al.</i> 1996
Grain hardness	Courtot x CS	187 DH	350 RFLP and SSRs	3	1A ^L , 5D ^S , 6D	75%	Perretant <i>et al.</i> 1999
Grain texture	NY18 x Clark's cream	78 F5 RILs	313 markers	4	1 on 5D ^S (XpinB)	>60%	Campbell <i>et al.</i> 1999
Starch viscosity	NY18 x Clark's cream	78 F5 RILs	181 markers (60% map coverage)	10	QTLs on: 1A ^S , 2A, 2B, 2D ^L , 3D ^L	22-44%	Udall <i>et al.</i> 1999
Starch pasting properties	Recombinant substitution lines	95 (4A) to 105 (7D)		1	2QTLs on 4A	24-56%	Araki <i>et al.</i> 2000
Flour colour	Schonburg x Yarlanka	150 RILs	217 RFLP + AFLP	4	1 major on 7A	41%	Parker <i>et al.</i> 1998
Milling yield	Schonburg x Yarlanka	150 RILs	217 RFLP + AFLP	4	3 on 3A, 5A, 7D	9-11%	Parker <i>et al.</i> 1999
Starch granule proportion	Barley: Steptoe x Morex	150 DH		1	1 QTL on 2H	13-17%	Borem <i>et al.</i> 1999
Soluble arabinoxylans	Synth x Opata, Courtot x CS	76 RILs 91 DH	>1000 markers 323 markers	1 1	1 major on 1B ^L	35-42% 32-37%	Martinant <i>et al.</i> 1998

CS: Chinese Spring

RILs: Recombinant Inbred Lines

BSA: Bulk Segregant Analysis

DH: Doubled Haploid lines

SSD: Single Seed Descent

Nb env.: number of environments

Among these traits, endosperm texture of bread wheat, most often referred to as grain softness versus hardness, has important technological consequences. Indeed grain softness rely to particle size index, i.e. the distribution of endosperm fragment size for a given energy used in the milling process. More importantly, for a given milling process, hard wheat flour have a higher proportion of damaged starch granules, which leads to a greater water absorption capacity at the dough making stage, thus indirectly increasing dough strength and tenacity. The genetic determinism of grain hardness/softness has been extensively studied (Fritz *et al.* 1995, Sourdille *et al.* 1996, Perretant *et al.* 1999, Campbell *et al.* 1999). A major QTL has been mapped at the extremity of chromosome 5D^S, which explain more than 60% of the variation in progenies from hard x soft wheat crosses, in the vicinity of genes puroindolin a and b (Gautier *et al.* 1994), although the direct implication of puroindolin in softness remains to be fully demonstrated (Giroux and Morris 1998). Other QTLs detected in hard x soft progenies were minor. However in a recent study of Recital x Renan , two medium hard wheat sharing the same ha allele, we found QTLs for grain hardness, which co-located with QTLs for grain protein content (Groos *et al.* 2001). Zanetti *et al.* (2001) reported co-locations between QTLs for hardness and grain protein content in their wheat by spelt derived population.

Starch properties do not depend solely from endosperm texture, but also from genuine characteristics of starch granules. One of these characteristics, related to end-use quality (although mostly for noodle-making), is starch viscosity, which relies to the ratio between amylose and amylopectin, the two types of glucose chain polymers from which starch molecules are made. QTLs for starch viscosity have been found in a progeny of a hard x soft cross, and mapped in 5 chromosome regions, including the three chromosomes of homeologous group 2, but not in those where the key-enzymes GBSS (granule bound starch synthase) are located, probably because the two parents were not polymorphic at these loci. In a study of recombinant substitution lines which segregated only for the homeologous regions bearing the GBSS loci, also named waxy Wx (i.e. 7A, 7D and 4A, which is translocated for the corresponding region of 7B), Araki *et al.* (2000) have confirmed the major effect of Wx-B1 gene (4A) on starch-pasting viscosity, but also identified a second QTL on 4A^S. To complete this survey of starch QTLs, let us mention the study of the barley progeny from Steptoe x Morex, in which Borem *et al.* (1999) reported a QTL for the proportion of starch granules (A versus B) on chromosome 2H.

Starch properties may also be greatly affected by the premature action of α -amylases when climatic conditions cause pre-harvest sprouting (PHS). Genetic tolerance to PHS has been studied in the progenies issued of two different crosses by Anderson *et al.* (1993), who found 4 QTLs in each progeny, which together explained 44 to 51% of the variation. In the progeny of a bread wheat x spelt wheat cross, Zanetti *et al.* (1999) reported 8

QTLs, the most important ones being located on chromosomes 5A(close to the Q locus for spike shape), 6D^L,7B^L(on which amy 1-3 are mapped). In none of these two studies, QTLs have been found in the vicinity of the R genes involved in red kernel colour, although a phenotypic relation between white colour and susceptibility to PHS has long been reported. In a recent study in the Recital x Renan progeny, Groos *et al.* (2001) have found QTLs for PHS on the homeologous group 3, close to the R genes, but also an additional QTL on chromosome 5A, which affected both PHS and grain colour.

Other important bread-making related traits are flour colour (possibly associated with flavonoid compounds of high nutritional value) and milling yield, a complex traits which depends on both kernel shape and endosperm-aleurone layer separability. These two traits have been studied in the same progeny by Parker *et al.* (1998, 1999). They reported a major QTL for flour colour on chromosome 7A, and 3 minor QTLs for milling yield.

Last but not least, a major QTL for water soluble arabinoxylan content has been reported in two reference bread wheat populations (Martinant *et al.* 1998). These components have important consequences on poultry feeding by reducing digestibility and increasing excrement moisture, but conversely they might have a protective effect on human health by reducing cholesterol absorption in digestive track. Their role in baking as a source of water release for gluten formation has also been suggested.

USEFULNESS OF QTL INFORMATION.

When considering all traits and populations studied so far, without taking into account redundant reports, all chromosomes have been identified to carry at least one QTL (Table 4).

Table 4: Summary of bread-making related QTLs in reports before end of 2000

Chromosome	A	B	D	H
1	10	10	3	2
2	6	4	5	2
3	4	4	2	0
4	4	2	1	0
5	6	4	4	1
6	3	4	2	1
7	2	3	3	2

The apparent under-representation of the D chromosomes may be simply a bias due to the lack of polymorphism known to occur on this genome with currently available markers and the lower map coverage compared to the A and B. Also chromosome 1A and 1B may be biased

upward, because some studies have given more weight to storage protein loci, supposed to be a priori good candidate genes for bread-making quality traits. It should be noticed however that these storage protein loci do not generally explained an important proportion of the quantitative variation, except in a study where one of the parental line was a very strong wheat (Khelifi and Branlard 1992). In a few other studies, genes with known functions such as α -amylases co-locate with QTLs and may thus be candidate genes. However the large majority of QTLs have still no candidate, and it could be a long way before identifying them. A first approach which can be anticipated for finding candidate genes is fine mapping, followed by chromosome landing using, for example, BAC libraries or mass mapping of Expressed sequence tags (EST) of genes expressed during grain filling.

However some, if not most of these reported QTLs may reveal to be false positive. Some studies have used a very limited number of progenies, and are thus subjected to sampling effect, leading to lack of detection power, associated with a high type-I error risk. For example, in a population of 50 individuals, false QTL with r^2 as high as 0.18 may be obtained by chance only in more than 5% of the sample (Charmet 2000). Thus reports of studies using less than 100 (or better 200) RILs must be taken with caution. Moreover, although some reports actually tried to keep an acceptable overall type-I error risk (i.e. of finding false positive), generally by using permutation tests, most of them still used standard lod-scores or pointwise probability threshold. It has well been demonstrated that such misuse of standard probability leads to many false positive in genome-wide scans (Lander and Kruglyak 1995). Instead much more stringent individual significance threshold must be used to keep the global false positive risk at acceptable level (e.g. 10^{-4} for human genome scan, Lander and Kruglyak 1995). Application of such proposed threshold would lead most QTLs that we have mentioned in tables 1, 2 and 3 to vanish.

Moreover, before putting huge effort in tentative cloning of QT-genes, one would worthwhile confirm its existence in a further sample, preferably in an independent experiment. This could be achieved, for example, by cross-validation, as suggested by Melchinger *et al.* (1998). Again, these authors found that many previously reported QTLs do not resist cross validation, and claim that we should only be confident in those QTLs, which reveal to be robust through resampling techniques such as cross-validation or bootstrap. In the study of Perretant *et al.* (2000), only 3 bootstrap-confirmed QTLs were presented for dough strength, out of 8 identified as significant in a first stage. Another way of confirming the true existence of QTLs is to design experiment to built isogenic material or to cumulate the favourable alleles in a single genotype and assess precisely the improvement for the trait of interest. Several methods have been proposed recently to achieve QTL pyramiding in inbred lines (van Berloo and Stam

1998, Charmet *et al.* 1999), which may be useful in bread-making research and improvement.

The consistency or stability of QTL effects across environment is another important factor to be considered. Most QTL studies reported in tables 1-3 used several environments, and authors generally stressed those QTLs which were detected in most of them. Detection, characterisation and use of QTLs which are specific to a given environment are still being worked out (e.g. Crossa *et al.* 1999).

Lastly, we should mention that most studies only dealt with additive, i.e. main effects of QTLs, which are the easiest to detect and to handle. Some studies did attempt to detect interactive QTLs, which may be expected to occur frequently, although the false positive risk is even higher than for main effect-QTLs. Perretant *et al.* (2000) mentioned such an interaction for dough strength, but they failed to confirm it through bootstrapping. Recently, Zanetti *et al.* (2001) reported 3 epistatic interactions between QTLs, which they had already detected for main effects. However detection of epistatic QTLs is not yet carried out systematically, probably because it requires more powerful designs and methods.

TOWARDS WHEAT GENOMICS

Starting from QTLs for cloning genes of bread-making quality is only one side of the problem. Recent large scale developments in molecular research have yielded other, more straightforward approaches. The most promising would be the construction of expressed sequences tags from cDNA libraries from a range of organs and development stages (e.g. grain during filling period). An international initiative (ITEC) is currently gathering sequence data from nearly twenty world contributors (ITEC 2001). Exploitation of such sequence information, and particularly comparison with model species annotated genomes would permit to assign putative functions to ESTs (Lazo *et al.* 1998). Thus, rather to look for genes behind the QTLs we shall have to look for QTLs to put forward predicted functional genes.

We will just give one example: One key-factor of bread-making quality probably resides in the relative quantity of gluten-forming proteins, namely HMW-glutenins, LMW-glutenins and gliadins. Understanding the regulatory system of their gene expression would thus be of primary importance. This involves research on the promoter region of coding genes (Anderson 1998), as well as the study of possible transcription factors (e. g. SPA, Albany *et al.* 1997). Sequence similarities have been found between ESTs from filling grains and the published sequence of SPA, indicating that this transcription factor is actually expressed at this stage.

THE PROMISE OF BIOTECHNOLOGIES

Once the functionality has been demonstrated either by association studies (e.g. Giroux *et al.*, 2000 for puroindoline and grain softness) or by genetic transformation (e.g. Krishnamurthy and Giroux 2001, again puroindoline/softness), the way is open to the genetic manipulation of these functional genes. This may be achieved by systematic assessment of allelic effects in genetic resources collections, followed by genetic transfer of the best alleles into adapted varieties (e.g. through marker assisted backcrosses), or by genetic engineering of either the coding sequence or the promoter region. One demonstrative example is the manipulation of HMW glutenin subunits in transgenic wheats. Barro *et al.* (1997) reported that improvement in dough functional properties was associated with the copy number of HMW-transgenes. Anderson *et al.* (1996) and Shimoni *et al.* (1997) have constructed synthetic and recombinant storage protein genes, which alter gluten polymer formation in transgenic wheat. Such experiments would provide new tools for quality improvement for a range of end-uses: baking, starch industry...

CONCLUSION

Wheat bread making quality still has an open field of research to be achieved in many disciplines. New achievement shall arise from the convergence of molecular biology, structural biochemistry and classical or quantitative genetics.

REFERENCES

- Albani D, Hammond-Kozack MCU, Smith C, Conlan S, Colot V, Holdworth M, Bevan MW, 1997. The wheat transcriptional activator SPA: a seed-specific bZip protein that recognizes GCN4-like motif in the bifactorial endosperm box of prolamine genes. *Plant Cell* 9:171-184.
- Anderson JA, Sorrells ME, Tanksley SD, 1993. RFLP analysis of genomic regions associated with resistance to pre-harvest sprouting in wheat. *Crop Sci.* 33 :453-459.
- Anderson OD, Kuhl J, Tam A, 1996. Construction and expression of a synthetic storage protein gene. 174:51-59.
- Anderson OD, Abraham-Pierce FA, Tam A, 1998. Conservation in wheat high-molecular-weight glutenin gene promoter sequences: comparisons among loci and among alleles of the glu-B1-1 locus. *Theor. Appl. Genet.* 96:568-576.
- Araki E, Miura H, Sawada S, 2000. Differential effects of the null alleles at the three Wx loci on the starch-pasting properties of wheat. *Theor. Appl. Genet.* 100:1113-1120.

- Barro F, Barcelo P, Rooke L, Tatham AS, Bekes F, Shewry PR, Lazzeri P, 1997. Transformation of wheat with high molecular weight subunit genes results in improved functional properties. *Nature Bio/Technology* 15:1295-1299.
- Bezant JH, Laurie DA, Pratchett N, Chojecki J, Kearsey MJ, 1997. Mapping of QTL controlling hot water extract and grain nitrogen content in a spring barley cross using marker regression. *Plant Breed.* 116 :141-145.
- Berloo R van, Stam P, 1998. Marker-assisted selection in autogamous RIL populations: a simulation study. *Theor. Appl. Genet.* 96 :147-154.
- Blanco A, de Laddomada C, Sciancalepore B, Simeone A, Devos KM, Gale MD, 1996. Quantitative trait loci influencing grain protein content in tetraploid wheats. *Plant Breed.* 115 :310-316.
- Blanco A, Bellomo MP, Lotti C, Pasqualone A, 1998. Mapping of quantitative trait loci for grain quality using recombinant inbred lines of durum wheat. *EWAC Newsletter* 1998 : 106-110.
- Blanco A, Bellomo MP, Lotti C, Maniglio T, Pasqualone A, Simeone R, Troccoli A, Di Fonzo N, 1998. Genetic mapping of sedimentation volume across environments using recombinant inbred lines of durum wheat. *Plant Breed.* 117:413-417.
- Borem A, Mather DE, Rasmusson DC, Fulcher RG, Hayes PM, 1999. Mapping quantitative trait loci for starch granule traits in barley. *J. Cereal Sci.* 29:153-160.
- Buschuk W 1998. Wheat breeding for end-product use. *Euphytica* 100:137-145.
- Carrillo JM, Rousset M, Qualset CO, Kasarda DD, 1990. Use of recombinant inbred lines of wheat for study of associations of high-molecular-weight glutenin subunit alleles to quantitative traits 1. Grain yield and quality prediction tests. *Theor. Appl. Genet.* 79:321-330.
- Campbell KG, Bergman CJ, Gualberto DG, Anderson JA, Giroux MJ, Hareland G, Fulcher RG, Sorrels ME, Finney PL, 1999. Quantitative trait loci associated with kernel traits in a soft x hard wheat cross. *Crop Sci.* 39 :1184-1195.
- Charmet G, 2000. Power and accuracy of QTL detection: simulation studies of one-QTL models. *Agronomie* 20:309-323.
- Charmet G, Robert N, Perretant MR, Gay G, Sourdille P, Groos C, Bernard S, Bernard M, 1999. Marker-assisted recurrent selection for cumulating additive and interactive QTLs in recombinant inbred lines. *Theor. Appl. Genet.* 99:1143-1148.
- Crossa J, Vargas M, van Eeuwijk FA, Jiang C, Edmeades GO, Hoisington D, 1999. Interpreting genotype x environment interaction in tropical maize using linked molecular markers and environmental covariables. *Theor. Appl. Genet.* 99:611-625.
- Elouafi I, Nachit MM, Elsaleh A, Asbati A, Mather DE, 2000. QTL mapping of genomic regions controlling gluten strength in durum (*Triticum turgidum* L. var durum). *Options Méditerranéennes. Series A, Séminaires méditerranéens* 2000, 40 :505-509
- Fritz AK, Cox TS, Gill BS, Sears RG, 1995. Marker-based analysis of quantitative traits in winter wheat x *Triticum tauschii* populations. *Crop Sci.* 35 :1695-1699.

Gautier MF, Aleman ME, Guirao A, Marion D, Joudrier P, 1994. Triticum aestivum puroindolines, two basic cystein-rich proteins: cDNA sequence analysis and developmental gene expression. *Plant Mol. Biol.* 25:43-57.

Giroux MJ, Morris CF, 1998. Wheat grain hardness results from conserved mutations in the friabilin components puroindoline A and B. *Proc. Natl. Acad. Sci. USA* 95:6262-6266.

Giroux MJ, Talbert L, Habernicht DK, Lanning S, Hemphill A, Martin JM, 2000. Association of puroindoline sequence type and grain hardness in hard red spring wheat. *Crop Sci.* 2000:370-374.

Groos C, Gay G, Perretant M-R, Gervais L, Bernard M, Dedryver F, Charmet G, *in press*. Study of the relationship between Pre-Harvest Sprouting and grain color by Quantitative Trait Loci analysis in a white x red grain bread wheat cross. *Theor. Appl. Genet.*

Groos C, Robert N, Bervas E, Charmet G., *submitted*. Analysis of genetic, environmental and genetic x environmental components for grain protein content, grain yield and thousand kernel weight in bread wheat. *Theor. Appl. Genet.*

ITEC, 2001. International Triticeae Mapping Initiative. *Proc. Plant and Animal Genome IX*, January 13-17, San Diego CA, USA.

Joppa LR, Du ChangHeng, Hart GE, Hareland GA, 1997. Mapping gene(s) for grain protein in tetraploid wheat (*Triticum turgidum* L.) using a population of recombinant inbred chromosome lines. *Crop Sci.* 37 :1586-1589.

Krishnamurthy K, Giroux MJ, 2001. Expression of wheat puroindoline genes in transgenic rice enhances grain softness. *Nature Biotech.* 19:162-166.

Khelifi D, Branlard G, 1992. The effects of HMW and LMW subunits of glutenin and of gliadins on the technological quality of progeny from four crosses between poor breadmaking quality and strong wheat cultivars. *J. Cereal Sci.* 16:195-209.

Kieffer R, Wieser H, Henderson MH, Graveland A, 1998. Correlations of the breadmaking performance of wheat flour with rheological measurements on a micro-scale. *J. cereal Sci.* 27:53-60.

Lander E, Kruglyak LK, 1995. Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. *Nature genet.* 11:241-247.

Lazo GR, Larka LA, Hsia CC, McCue KF, Sorrels M, Matthews DE, Au M, Federspiel NA, Anderson OD, 1998. Assigning putative gene functions to mapped probe loci in the GrainGenes genome database and sequencing of wheat endosperm cDNAs. *Proc. Plant and Animal Genome VI Conf*, San Diego, CA, USA, 18-22 January.

Manifesto MM, Feingold S, Hopp HE, Schlatter AR, Dubcovsky J, 1998. Molecular markers associated with differences in bread-making quality in a cross between bread wheat cultivars with the same high Mr glutenins. *J. cereal Sci.* 27:217-227.

Martinant JP, Cadalen T, Billot A, Chartier S, Leroy P, Bernard M, Saulnier L, Branlard G, 1998. Genetic analysis of water-extractable arabinoxylans in bread wheat endosperm. *Theor. Appl. Genet.* 97 :1069-1075.

- Melchinger AE, Utz HF, Schön CC, 1998. Quantitative trait locus (QTL) mapping using different testers and independent population samples in maize reveals low power of QTL detection and large bias in estimates of QTL effects. *Genetics* 149:383-403.
- Morris CF, 1998. Genetic determinants of wheat grain quality. *Proc. Ninth Intern. Wheat genet. Symp., Saskatoon, Canada, 2-7 August 1998* : 245-253.
- Oury FX, Chiron H, Pichon M, Giraud A, Berard P, Faye A, Brancourt-Hulmel M, Rousset M, 1999. Reliability of indirect selection in determining the quality of bread wheat for French bread-making. *Agronomie* 19:621-634.
- Parker GD, Chalmers KJ, Rathjen AJ, Langridge P, 1998. Mapping loci associated with flour colour in wheat (*Triticum aestivum* L.) *Theor. Appl. Genet.* 97:238-245.
- Parker GD, Chalmers KJ, Rathjen AJ, Langridge P, 1999. Mapping loci associated with milling yield in wheat (*Triticum aestivum* L.) *Mol. Breed.* 5:561-568.
- Payne PI, 1987. Genetics of wheat storage proteins and the effect of allelic variation on breadmaking quality. *Ann. Rev. Of Plant Physiol.* 38:141-153.
- Perretant MR, Cadalen T, Charmet G, Sourdille P, Nicolas P, Bœuf C, Tixier MH, Branlard G, Bernard S, Bernard M, 2000. QTL analysis of bread-making quality in wheat using a doubled haploid population. *Theor. Appl. Genet.* 100 :1167-1175.
- Prasad M, Varshney RK, Kumar A, Baylan HS, Sharma PC, Edwards KJ, Singh H, Dhaliwal HS, Roy JK, Gupta PK, 1999. A microsatellite marker associated with a QTL for grain protein content on chromosome arm 2DL of bread wheat. *Theor. Appl. Genet.* 99 :341-345.
- Rousset M, Brabant P, Kota RS, Dubcovsky J, Dvorak J, 2001. Use of recombinant substitution lines for gene mapping and QTL analysis of breadmaking quality. *Proc. 6th IntreN; Wheat Conference, Budapest 5-9 June 2000, Kluwer Acad. Pub. (in press).*
- Sax K, 1923. The association of size differences with seed-coat pattern and pigmentation in *Phaseolus vulgaris*. *Genetics* 8:552-560.
- Shimoni Y, Blechl A, Anderson OD, Galili G, 1997. A recombinant protein of two HMW glutenins alter gluten polymer formation in transgenic wheat. *J. Biol. Chem.* 272:15488-15496.
- Snape JW, Laurie DA, Quarrie SA, 1995. Comparative QTL analysis and its application to cereal breeding. *Proc. IAEA « Induced mutations and molecular techniques for crop improvement », Vienna, Austria 19-23 June 1995* : 39-49
- Sourdille P, Perretant MR, Charmet G, Leroy P, Gautier MF, Joudrier P, Nelson JC, Sorrells ME, Bernard M, 1996. Linkage between RFLP markers and genes affecting kernel hardness in wheat. *Theor. Appl. Genet.* 93:580-586.
- Udall JA, Souza E, Anderson E, Sorrells ME, Zemetra RS, 1999. Quantitative trait loci for flour viscosity in winter wheat. *Crop Sci.* 39 :238-242.
- Zanetti S, Keller M, Winzeler M, Saurer W, Keller B, Messmer M, 1998. QTL for quality parameters for bread-making in a segregating wheat by spelt population. *Proc. Ninth Intern. Wheat genet. Symp., Saskatoon, Canada, 2-7 August 1998* : 273-276.

Zanetti S, Keller M, Winzeler M, Saurer W, Keller B, Messmer M, 1999. QTL for quality parameters for bread-making in a segregating wheat by spelt population. In : « Genetics and breeding for quality and resistance », Proc. XV EUCARPIA congress, Viterbo, Italy, 20-25 September 1998. Kluwer Acad. Pub., Dordrecht, Germany :357-360.

Zanetti S, Winzeler M, Feuillet C, Keller B, Messmer M, 2001. Genetic analysis of bread-making quality in wheat and spelt. *Plant Breed.* 120:13-19.